
Proceedings
Investigation of the integration of the weight function already has the potential to improve the performance of the detection methods. Further processing can be performed based on the knowledge of the position of these pixels and the search for the staff lines could be guided through them. Afterwards, we extend the strong staff-pixel definition for grayscale images.

3 Strong Staff Pixels

We start by introducing the concept of strong staff-pixel in binary images. The integration of this concept in the weight function already has the potential to improve the performance of the detection methods. Further processing can be performed based on the knowledge of the position of these pixels and the search for the staff lines could be guided through them. Afterwards, we extend the strong staff-pixel definition for grayscale images.

3.1 Strong Staff-Pixels

Intuitively, a strong staff-pixel is a pixel with a high probability of belonging to a staff line. The decision of accepting/rejecting a pixel as a strong staff-pixel is not made independently for each pixel but builds on the length encoding as previously mentioned. All pixels in the same run will either be considered as staff pixels or not staff pixels. The decision approach consists of scanning the run-length encoding of each column of the binary image of the music sheet in order to find black runs of staffline_height pixels followed or preceded by a white run of staffspace_height pixels. See Fig. 1 for an example. The pixels in the black runs that meet this condition form the set of Strong Staff-pixels (SSP).

In the grayscale domain, instead of computing the previously explained method for just only one binary image obtained by a state-of-the-art binarization technique, the scanning RLE is performed for every possible binary image by varying the threshold from a low to a high limit. In this manner, the final set of SSP will be formed by accumulating sets of black runs with a height equal to staffline_height that are followed or preceded by a white run of staffspace_height pixels.

3.2 Improved Weight Function in the Binary Domain

The initial version of the weight function was presented in [2]. Here we propose the use of the concept of SSP to penalize or to favor the current pixel. Hence, if the current pixel belongs to the set of SSP (SSP = 1) then a benefit term is incorporated to the weight. Otherwise, the black pixels (pixelValue = 0) that do not belong to this group (SSP = 0) and belong to a vertical run of black pixels higher than the staffspace_height are penalized.

4 Grayscale Staff Line Detection

For binary staffs, it was sufficient to assign a low cost for incident black pixels and a high cost for the white pixels. Now, instead of only...
The pixels with high probability of belonging to staff lines are represented with red color.

Figure 1: An illustrative example of the strong staff-pixel algorithm. The staffspace_height is 3 and staffline_height is 2.

values (\((0,1]\), background and foreground) we have a range of values in \([0,1]\). Therefore, the weight function must be generalized to this new domain.

Our proposed weight function is based on a sigmoid function where the parameters are chosen to favor the lumiance levels of staff — see Eq. 1. The rationale is to extend the penalization of pixels in the binary procedure that can be seen as a step function (pixels binarized to white are assigned a high cost; pixels binarized to black are assigned a low cost). With this in mind, the sigmoid function selected for 4-neighbourhoods was the following:

\[
w(p) = b/(1 + \exp(-(p - \alpha)/\beta)) + a
\]

where \(p\) is the minimum of the values of the two pixels incident with the edge. In the case of 8-neighbourhoods the weight is \(\sqrt{2}\) times that value. The \(a\) and \(b\) are normalization factors for the values range of our cost function and they were both set to 4; \(\alpha\) was set as the threshold value from the BLIST binarization [3].

The \(\beta\) parameter is inversely related to the slope of the sigmoid at the center of symmetry \(\alpha\). Since different scores can have different dispersion of values within the staff lines, one chose to set the \(\beta\) proportional to the standard deviation, \(staff\text{value}_{\text{std}}\), of the gray intensity levels of the SSP’s. Experimentally \(\beta\) was set to 0.8 \times \(staff\text{value}_{\text{std}}\). The pseudo-code for the weight function is provided in Listing 1.

```c
WeightFunction(pixelValue1, pixelValue2, a, b, alpha, beta,
    vRun1, vRun2, nearestVRun1, nearestVRun2, NeighbourhoodType)
{
    value = min(pixelValue1, pixelValue2);
    weight = sigmoidFc(value,a,b,alpha,beta,NeighbourhoodType);
    if( vRun1<=STAFFLINEHEIGHT )
        OR(vRun2<=STAFFLINEHEIGHT) AND value <= alpha)
        weight = weight - delta;
    if( nearestVRun1=STAFFSPACEHEIGHT=STAFFLINEHEIGHT)
        OR(nearestVRun2=STAFFSPACEHEIGHT=STAFFLINEHEIGHT)
        weight = weight + delta;
    return weight;
}
```

Listing 1: Pseudo-code for the weight function. The sigmoidFc is the Equation 1. The \(delta\) penalizing term in the weight function was set to 1. For efficiency, weights were designed with integer values.

5 Metrics and Results

In this work two different evaluations were carried out: i) the quality of the pixels that belongs to the set of SSP and ii) the quality of the line detection algorithm. The evaluation of the quality of the SSP was obtained by comparing the resulting image with the reference positions of symbols and staff lines. The evaluation of the performance of the line detection algorithm was done through the evaluation of the performance of the staff line detection and removal algorithms. A set of 76 of handwritten music scores, for which reference staff lines were manually outlined, was used.

5.1 SSP Evaluation

For the first experimental testing the resulting images with only staff line pixels were compared with images with only true staff lines pixels to count the number of false pixels detected. This number restricts the quality of the SSP. The number of missed pixels in this situation is not relevant provided that we have pixels in real positions of staff lines. The staff lines ground truth images are composed by the true positions of the staff lines with breaks on the symbols positions. The symbols ground truth images composed by music symbols without noise were compared to the resulting images with only staff-pixels. The aim was to check if the false positives (pixels wrongly classified as staff lines) detected by the algorithm belong to noise or to music symbols.

The number of pixels wrongly classified as non-staff lines was 7% (4% as symbols and 3% as noise) against 93% for pixels classified as staff lines, which represents a very satisfactory result. Usually, in a music score, most of the symbols appear on the staff lines and noise could be anywhere. As the goal is to follow the staff lines as closely as possible it is better to classify staff-pixels as symbols than as noise.

5.2 Staff Line Detection Evaluation

For the evaluation of the staff line detection algorithm two metrics were considered: the number of unmatched detected staff lines (false positives) and the number of unmatched reference staff lines (missed detections).

The metrics were computed by the average Euclidian distance between each reference staff line and each actually detected staff line, where the matching problem on the resulting bipartite graph was solved by minimizing the distance [2]. Both binary and grayscale domains were compared. The SSP approach for binary images was also compared with the first weight function proposed in [1]. The binarization of the images was carried out using the BLIST method [3]. The outcome of the algorithms can be seen in Table 1. When the SSP was used to guide the searching of the staff lines the results were better. In the case of grayscale images the staff line detection algorithm obtained better results.

<table>
<thead>
<tr>
<th>Weight function</th>
<th>False detection rate</th>
<th>Miss detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLIST</td>
<td>1.0 (2.0)</td>
<td>1.2 (2.3)</td>
</tr>
<tr>
<td>BLIST</td>
<td>0.8 (2.1)</td>
<td>0.9 (2.2)</td>
</tr>
<tr>
<td>Grayscale Sigmoid</td>
<td>0.7 (2.2)</td>
<td>0.9 (2.8)</td>
</tr>
<tr>
<td>Grayscale Sigmoid</td>
<td>0.7 (1.8)</td>
<td>0.8 (2.0)</td>
</tr>
</tbody>
</table>

Table 1: Detection performance in percentage: mean (standard deviation).

6 Conclusion

In this paper, the concept of SSP was introduced in order to guide the staff line detection algorithm proposed in [2]. The new methodology improves the results of binary based detection and allows generalization of the framework to the grayscale domain.

7 Acknowledgement

This work is financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the PCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281.

References

Cancer cell tracking using a Kalman filter

1. Introduction

The automatic analysis of cancer cell’s mobility based on time-lapse microscopy images has gained relevance due to the increase in data amount that biology researchers analyze [7]. Usually, for studying cancer cell mobility, cancer cells are placed on top of uncoated surfaces or on surfaces coated with extracellular matrix components and a time lapse video is collected as interaction with the substrate occurs.

Cell mobility analysis is the study of cell’s positions variation through time, through automatic image tracking of cells. Cell tracking can be performed by detection-association tracking approaches [1] or by probabilistic state space modeling [6]. Using cell detection-association based tracking, cells are first detected in each frame and cell to cell associations are performed between adjacent frames based on spatial and descriptor similarity. The tracking process fails when a cell is not detected. Considering motion models there are techniques such as Kalman filters where some prior knowledge on the nature of cell motion is assumed and applied to estimate the location of cells prior to detection. Kalman filter was applied by Huth et al. for the automatic cell tracking in DIC microscopy image sequences [3]. Li et al. also used this filter to perform online tracking of migrating and proliferating cells [4].

In our work we propose to perform cell tracking using Kalman filtering comparing the assumption of a random or a constant velocity motion and using LoG filter based cell detection with cross correlation coefficient or SIFT descriptors for cell characterization.

2. Methodology

To perform automatic cancer cell’s mobility analysis we use a Kalman filter. We perform the tracking of each cell independently using the Laplacian of Gaussian filter to detect cells and considering SIFT and cross correlation coefficients for cell characterization. We tested the assumption of random and constant velocity motion models.

2.1. Cell tracking using a Kalman filter

The Kalman filter performs state estimation of a dynamic system from a series of incomplete or noisy measurements minimizing the mean of the squared error and it consists of two alternating operations: first, it predicts the system’s state and its uncertainty and then combines that prediction with the noisy measurement in order to correct the initial prediction. The system and the probability density function at each state follows a Gaussian distribution and the equations for the Kalman filter consist in: time update or prediction and measurement update or correction:

\[
\begin{align*}
\text{Prediction:} & \quad X_k^* = AX_{k-1} + w_{k-1}, \\
\text{Measurement:} & \quad Z_k = HX_k + v_k.
\end{align*}
\]

Where \( X_k^* \) is the predicted state vector at frame \( k \), \( A \) is the dynamics matrix which relates the state at frame \( k-1 \) with the state in the current frame \( k \). The prediction and measurement Gaussian noise are represented by \( w_{k-1} \) and \( v_k \), respectively, and \( H \) is the measurement matrix that relates the state vector to the measurement \( Z_k \).

We define both random position and constant velocity motion dynamics. This leads to different dynamics matrices \((A^1, A^2)\) and also state variables:

- Random motion: assumes that the next state is defined by the previous state considering random motion and SIFT descriptors.

\[
A^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad X_k = [x_k, y_k, radius_k]^T.
\]

- Constant velocity motion: makes the assumption that the next state is given by the previous state plus an addition of the cell velocity in the previous frame:

\[
A^2 = \begin{bmatrix} 1 & dt & 0 & 0 \\ 0 & 1 & dt & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad X_k = [x_k, y_k, vx_k, vy_k, radius_k]^T.
\]

For the task of cell detection we use the LoG filter which is based on the image scale-space representation to enhance the blob like structure as introduced by Lindeberg [5]. We set the scale of the filter given the expected range of the cell radius, and we perform detection of cells by detecting local maxima of LoG response in the input image (figure 1(a)). The detected maxima enable us to estimate the position and radius of cells (figure 1(b)) [2].

Based on the detected cells, we compute, for each one, cell descriptors that together with the predicted state \( X_k^* \) are used to obtain the measurement \( Z_k \). This is performed by finding the most similar cell both in spatial distance (using \( X_k^* \)) and descriptors similarity. The use of cell descriptors enables the characterization of cells across frames. We use pixel information with cross correlation coefficient (ccc) or SIFT descriptors to obtain such characterization by matching the cell detection corresponding to the previous measurement \( Z_{k-1} \) with the detected cells in frame \( k \). The template size in the cross correlation method and the scale assumed in the SIFT descriptor are given by the LoG detected radius of each cell.
Table 1: Performance results for cancer cell tracking using Kalman filter and feature descriptors (cross correlation coefficient (ccc) and SIFT)). Values are presented in percentage.

<table>
<thead>
<tr>
<th>Similarity measure</th>
<th>Constant position</th>
<th>Constant velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detection Accuracy</td>
<td>Correct</td>
</tr>
<tr>
<td>ccc SIFT</td>
<td>73.1</td>
<td>63.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We use $Z_k$ to correct the initial prediction considering the Kalman gain ($K$), obtaining the final state vector $X_k$ according to:

$$X_k = X_k^\phi + K(Z_k - HX_k^\phi).$$  \hspace{1cm} (3)

The state vector $X_k$ of each cell in each frame allow us to obtain the track of each cell along the video sequence.

2.2 Results and Discussion

To evaluate the performance of the proposed tracking approach, we tested on a time-lapse video containing 156 brightfield frames considering 30 cells randomly selected. The automatic tracking of each cell was performed individually.

In Table 1 we present the performance results where we observe that better results were obtained considering SIFT descriptors for cell shape characterization and assuming a random motion. In figure 1 (c) it is possible to observe the final result of the tracking process.

3 Software prototype

In order to allow biology researchers to use our work in their experimental analysis tasks we developed an easy to use software to perform cell detection and tracking. In Figure 2 is visible the graphical user interface of the developed software application where the tracking result of several cells is presented.

In addition to facilitate cell annotation by biologist researchers this will help us in the validation of our system. Currently this validation is performed by comparing the obtained tracks from the automatic approach with the tracks obtained manually by experts. By using this software the obtained tracks are corrected on-line. For the biologist researchers it is an easy task and they obtain from the software what they need for cell analysis and for our work we obtain tracks validated that we can use for methodology validation.

4 Conclusion

We proposed an automatic approach for cancer cells tracking in brightfield microscopy images in order to analyse their mobility. For that a Kalman filter was used comparing between constant position and constant velocity motion models. Results showed that the assumption of a random motion and the use of SIFT descriptors for the measurement process led to better tracking results.

As future work we aim at tracking the complete cell population using a model that includes the model of shape and mobility together.
Abstract
Evaluating parasite infection indexes on in vitro cell cultures is a practice commonly employed by biomedical researchers to address biological questions or to test the efficacy of novel anti-parasitic compounds. In the case of Leishmania infantum, infection indexes are usually determined either by visual inspection of cells directly under the microscope or by counting digital images using appropriate software. In either case assessment of infection indexes is time consuming, thus motivating the creation of automatic image analysis approaches. Our aim is to develop a fully automatic methodology for infection indexes evaluation. In our previous work we approach detection problem with a feedback loop that tunes to the size of the object of interest, based on the DoG filter. We propose using linear spectral unmixing in our cell detection loop to improve image quality for higher analysis performance. Our linear spectral unmixing approach can improve image quality and the final detection results, particularly when the image being processed presents overlapping spectral profiles.

1 Introduction
Leishmania infantum is a unicellular parasite that causes human and canine leishmaniasis in countries of the Mediterranean basin, Portugal included. In its mammalian hosts, Leishmania infects and multiplies inside macrophages. One strategy that is commonly employed to study the mammalian stage of Leishmania makes use of in vitro macrophages cultures infected with this parasite. Under this setting, macrophage infection indexes are used as a measure of parasite growth and survival. Infection indexes are determined by multiplying the percentage of infected macrophages by the average number of Leishmania per cell. Both these parameters are determined by manual counting of cells which is time consuming and subjective [2]. This motivates the creation of automatic cell culture analysis approaches. Most approaches for cell detection in fluorescence microscopy images are based on automatic image segmentation [3, 8, 10]. However, such approaches are governed by parameters that are not robust to image quality and are complex for someone unfamiliar with image processing, requiring experts for readjustments, reducing robustness and usability [9, 10].

Local image filters have been introduced to aid in cell detection based on shape. Usay et al. proposed the use of the Laplacian of Gaussians for cell detection, given their approximated circular shape [7]. Esteves et al. proposed the use of local convergence filters for the detection and shape estimation of cell nuclei [4]. In our previous work we proposed a fully automatic system for assessment of infection index on macrophage cell cultures based on feedback loop that iteratively detect the size of cellular components [5]. However, some cell cultures images suffer from fluorescence cross-bleed, making analysis more error prone. In typical use of fluorescence microscopy, either the sample to be observed is tagged with fluorescent molecules (fluorophores) or the sample itself has fluorescing properties. When excited these molecules release photons, which can be recorded as fluorescent intensity resulting in a spectral profile particular to each fluorophore. However, it is common that spectral profiles of different fluorophores overlap. Spectral imaging coupled with image analysis using linear spectral unmixing can be employed to segregate mixed fluorescent signals and more clearly resolve the spatial contribution of each fluorophore, facilitating image segmentation. Al-Kofahi et al. proposed the use of multi-scale Laplacian-of-Gaussian filtering on histopathology images with unmixed fluorophores spectral profiles to automatically detect cell nuclei [1].

2 Methods
With the aim of developing a fully automatic methodology for evaluation of infection indexes in in vitro cell cultures, we approach cell detection with a feedback loop that tunes the size of the object of interest, based on DoG filter [5]. We use the location of the three detected cell areas, corresponding to the cellular components, as inputs for spectral unmixing. Using linear spectral unmixing we improve image quality for cell detection and for visualization. As final goal, we calculate infection indexes by analyzing the overlap between the different segmented areas of interest.

2.1 Localization of cell structures
Localization of cell structures (nuclei, parasites, cytoplasm) is obtained using two different approaches. Nuclei and parasites are detected using a DoG local filter. The cytoplasm positions are gathered using a watershed. Figure 1 shows an example of cell structure localization using these approaches.

Prior to the detection we perform background subtraction in the green and blue channels and in the red channel we do adaptive histogram equalization.

2.1.1 Nuclei and parasites detection
To find the nuclei and parasite locations we take advantage of the scale selection properties of the Difference-of-Gaussians (DoG) filter, which is an approximation of the Laplacian-of-Gaussian filter [6]. This filter can be defined as the subtraction between two Gaussian smoothed copies of the original image:

\[ D(x, y, \sigma) = L(x, y, \sigma) - L(x, y, \sigma + k \sigma) \]
where \( k \) is the difference in the standard deviation of the Gaussians and \( L \) is a Gaussian smoothed image defined by:

\[
L(x, y, \sigma) = g(x, y, \sigma) * I(x, y)
\]

where:

\[
g(x, y, \sigma) = e^{-\frac{(x^2 + y^2)}{2\sigma}}
\]  

(2)

The filter response is higher for objects within the DoG filter’s size band. To overcome the need to know the size of the objects, DoG parameters are tuned based on the data. Starting with a initial size estimate, we use a feedback loop where the assumed object size in each iteration is a weighted sum between the old size and the new size estimate. A verification if the detected objects size is coherent with that initial guess is performed, if not the system updates the size estimate repeating the process. We obtain the detections using Otsu’s automatic thresholding on the filter response image. To further increase robustness we do not fully accept all detections, instead we fit a Gaussian to the detected objects’ size data and remove the top and bottom 5% of all detections (both are ignored for size estimation purposes). For the final image result the smaller 5% detections are considered noise and are removed, and the top 5% are split based on regional maximum existing in the DoG filtered image from the last iteration of the feedback loop.

2.1.2 Cytoplasm segmentation

While both macrophage nuclei and parasite detection can be performed based on their overall shape, macrophage’s cytoplasm has no characteristics shape. The only prior knowledge we can use for cytoplasm segmentation is that the cell nucleus is inside the cytoplasm. Given the location of nuclei, we use seeded watershed as our cytoplasm segmentation approach, using the nuclei location as the seed for the watershed [10].

2.2 Automatic supervised linear spectral unmixing

In this work we implement supervised linear spectral unmixing in an automatic way to isolate the spectrum belonging to each one of three cellular structures. One drawback of using supervised linear spectral unmixing is that the cell nucleus is inside the cytoplasm. Given the location of nuclei, we use seeded watershed as our cytoplasm segmentation approach, using the nuclei location as the seed for the watershed [10].

To evaluate our methods performance we calculated the detections errors (DE) for nuclei and parasites. Since the purpose of this work is to automatically estimate the infection level in the cell culture we calculated the infected cell count error (ICE) and the number of parasites infecting each cell error (NPIE). The detections errors are obtained comparing the automatic results with the groundtruth.

In Table 1 are the detection errors for our proposed method, with and without the supervised linear unmixing step. Regarding the infection level estimation, the main objective of our approach, the infected cell count error was 3.3% with a number of parasites infecting each cell error of 14.7%. While for overall images there was little gain from using spectral unmixing (non statistically significant), there are considerable improvements for images with fluorescence cross-bleeding problems.

Table 1: Results comparing cell detection and parasite infection with and without linear spectral unmixing for the full 86 image dataset (FD) and for a dataset composed of 10 images exhibiting fluorescence cross-bleed problems (CD).

<table>
<thead>
<tr>
<th>Method</th>
<th>Infection Index</th>
<th>Nuclei DE(%)</th>
<th>Parasites DE(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FD</td>
<td>ICE(%)</td>
<td>NPIE(%)</td>
<td>DE(%)</td>
</tr>
<tr>
<td>Normal</td>
<td>3.6</td>
<td>14.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Unmixing</td>
<td>3.3</td>
<td>14.7</td>
<td>4.9</td>
</tr>
<tr>
<td>CD</td>
<td>ICE(%)</td>
<td>NPIE(%)</td>
<td>DE(%)</td>
</tr>
<tr>
<td>Normal</td>
<td>2.7</td>
<td>22.9</td>
<td>4.3</td>
</tr>
<tr>
<td>Unmixing</td>
<td>2.9</td>
<td>5.9</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Table 2: Detection results for the full 86 image dataset (FD) and for a dataset composed of 10 images exhibiting fluorescence cross-bleed problems (CD).

After performing linear spectral unmixing and luminance regularization, the output image is again fed to the detection algorithm described in section 2.1.

2.3 Parasite infection index evaluation

A macrophage is considered infected when parasites are inside its cytoplasm, this translates to an overlap between parasites and cytoplasmic image area. Given that we already obtained both image areas’ segmentation, we check, for each macrophage if there are parasites in its cytoplasmic area, using their central location. The final measures for the evaluation of the parasites’ infection index are the number of macropages in an image that are being attacked by parasites (parasites overlap with that macrophage’s cytoplasm) and the number of parasites infecting each macrophage (number of parasites inside the cytoplasm of the infected macrophage).

3 Results

To evaluate our methods performance we calculated the detections errors (DE) for nuclei and parasites. Since the purpose of this work is to automatically estimate the infection level in the cell culture we calculated the infected cell count error (ICE) and the number of parasites infecting each cell error (NPIE). The detections errors are obtained comparing the automatic results with the groundtruth.

In Table 1 are the detection errors for our proposed method, with and without the supervised linear unmixing step.

To evaluate our methods performance we calculated the detections errors (DE) for nuclei and parasites. Since the purpose of this work is to automatically estimate the infection level in the cell culture we calculated the infected cell count error (ICE) and the number of parasites infecting each cell error (NPIE). The detections errors are obtained comparing the automatic results with the groundtruth.

In Table 1 are the detection errors for our proposed method, with and without the supervised linear unmixing step.

Regarding the infection level estimation, the main objective of our approach, the infected cell count error was 3.3% with a number of parasites infecting each cell error of 14.7%. While for overall images there was little gain from using spectral unmixing (non statistically significant), there are considerable improvements for images with fluorescence cross-bleeding problems.

The spectral unmixing step allows a better detection of the macrophages cell limits (see Figure 2) resulting in a correct overlap between parasites and infected cells, bringing the results closer to the groundtruth.

4 Conclusions

Our proposed approach for performing linear spectral unmixing, with a set of points collected from the detected cellular regions, was able to improve image quality and the final detection results, particularly when the image being processed presents overlapping spectral profiles.

Results obtained showed that our approach has a 3.3% error comparing with the manual annotated results for infection level estimation and within 5% and 24% for nuclei and parasite detection respectively. Future work will be focused on the adaptation of the detection parameters for each experiment, instead of each image, as imaging condition do not change between images of the same culture. It is also important to find a suitable measure to automatically decide if a particular image presents scrambled spectral profiles and needs to be processed by linear unmixing.
Acknowledgements

This work was financed by FEDER funds through the Programa Operacional Factores de Competitividade COMPETE and by Portuguese funds through FCT Fundação para a Ciência e a Tecnologia in the framework of the project PEst-C/SAU/LA0002/2011. PQ is a Ciência2008 awardee.

References


Abstract

The present work aimed to provide a robust algorithm for biometric identification, based on the Viola-Jones algorithm for the segmentation of specific facial regions, and posterior feature extraction using SIFT keypoint descriptors. We trained M-order Gaussian Mixture Models to represent the feature probability distribution of each individual. Experimental results on the MobBIO multimodal database show that the proposed solution is able to match 68.1% of the images to the respective individual, reaching peak performance for M = 16 and the region corresponding to both eyes.

1 Introduction

Several human biological traits show a considerable inter-individual variability. Biometrics works by recognizing patterns within these biological traits to perform accurate recognition. The growing need for reliability and robustness, raised some expectations and became the focal point of attention for research works on biometrics [1].

Over the past few years face recognition has been on the spotlight of many research works in biometrics. The face is an easily acquirable trait with a high degree of uniqueness. These marked advantages, however, fall short when low-quality images are presented to the system. Several recent works have tried to explore alternative hypothesis to face this problem, either by developing more robust algorithms or by exploring new traits to allow or aid in the recognition process [7].

On the present work we explore the possibility of performing biometric recognition on reduced regions-of-interest inside the face, such as the eyes, the nose or the mouth. By working on smaller regions we attempt to offer an alternative to face recognition when only partial data is made available or when heterogeneous conditions characterize variable parts of the image. We also expect a smaller computational complexity and, thus, faster processing times as a result of the aforementioned process.

2 Methodology

The proposed algorithm aimed to perform biometric recognition on conceptually well-defined regions-of-interest in face image. Each image is processed using the Viola-Jones algorithm [6], yielding a segmentation of each individual trait of interest: mouth, nose and eyes.

After each region’s segmentation, the Scale Invariant Feature Transform (SIFT) strategy is used to compute a set of scale and rotation invariant keypoint descriptors to achieve a meaningful representation of each segment’s content. Given a dataset of N individuals, each individual’s data – i.e. set of SIFT keypoints – is modelled using Gaussian Mixture Models (GMM), an unsupervised learning procedure that yields a probability distribution of observations describing each individual’s specific variations of the tested biometric trait. Such models are capable of capturing the empirical probability density function (PDF) of a given set of feature vectors, so as to faithfully model their intrinsic statistical properties [5].

The choice of GMM to model feature distributions in biometric data is extensively motivated in many works of related areas. From the most common interpretations, GMMs are seen as capable of representing broad “hidden” classes, reflective of the unique structural arrangements observed in the analysed biometric traits [5]. Besides this assumption, Gaussian mixtures display both the robustness of parametric unimodal Gaussian density estimates, as well as the ability of non-parametric models to fit non-Gaussian data [4]. This duality, alongside the fact that GMM have the noteworthy strength of generating smooth parametric densities, confer them a strong advantage as generative model of choice for modelling.

All models are trained on sets of 128-dimensional Scale Invariant Feature Transform (SIFT) keypoint descriptors, \( x_{i=1} \ldots x_{i=N} \), extracted from the images of a specific individual, \( i \). Such features present invariance to image scaling, translation, rotation and partial invariance to illumination changes and affine or 3D projection [2], conferring them strong appeal in the area of unconstrained biometrics.

The trained models are validated and the global performance of the system assessed by the projection of the test data, \( x_{test} = \{x_{1} \ldots x_{N}\} \), onto all of the \( N \) previously trained individual-specific models. Decision is performed by a maximum likelihood test, using the average maximum likelihood value for all vectors \( x_{i}, \forall i \in \{1..N\} \). The global pipeline of the training and test phases of the proposed algorithm is schematized in Figure 1.

![Figure 1: Flow chart of the enrollment (model training) and identification (model validation) tasks of the proposed algorithm.](image)

3 Results and Discussion

3.1 Tested dataset

The proposed algorithm was tested on the MobBIO multimodal database [3], created in the scope of the 1st Biometric Recognition with Portable Devices Competition 2013, integrated in the ICIAR 2013 conference (https://www.iciar.uwaterloo.ca/iciar13/). The main goal of the competition was to compare different methodologies for biometric recognition using data acquired with portable devices. We tested our algorithm on the face modality present on this database. Regarding this modality, the database is composed by 105 individuals and a total of 2658 images. The train dataset is composed by 8 face images, while the test dataset presents a total of 1496 images with variable number of images per individual. Figure 2 depicts some examples of such images.

![Figure 2: Examples of face images in the MobBIO database.](image)
3.2 Recognition performance

On this section we discuss the most relevant results. Concerning the detection of the relevant regions-of-interest (ROI), the Viola-Jones algorithm’s failure-to-enroll (FTE) ratio can be defined as the ratio of ROI of each category – eyes, mouths and noses – that were not correctly detected. Table 1 presents the FTE results for each category. It can be noted that the eyes present the highest FTE, probably due to the more constrained nature of images used during the Haar cascade training phase (for example no glasses and constant illumination). Figure 3 depicts some examples of correctly detected structures of interest.

![Figure 3: Examples of detected regions-of-interest in some face images from the MobBIO database.](image)

Regarding the order of the GMM, \( M \), we observed a performance peak at \( M = 16 \). This tendency can be readily observed by the analysis of Figure 4. Working with higher orders could also result in some loss of generality and overfitting to noise, as well as slower testing time. For these reasons we present all further results with respect to the aforementioned value of the \( M \) parameter (see Table 1).

![Figure 4: Rank-1 recognition rates for variable values of the GMM order parameter \( M \).](image)

Another probable explanation for the performance degradation when working with higher orders might be connected to the stop criteria in the EM training algorithm: if a fixed number of iterations is defined, low order models will probably end in a better fit than more complex ones. This results in a stronger modelling of the data when working with low orders.

Table 1 summarizes the most significant results obtained for the optimal performance parameter of \( M = 16 \). We analyse the rank-1 recognition rate, that is, the ratio of correctly identified individuals in a 1:N comparison. In this setup each test image is compared against every one of the \( N \) individual-specific models, trained in the enrollment phase, and the model that outputs the highest likelihood is associated with the detected identity.

<table>
<thead>
<tr>
<th></th>
<th>Mouth</th>
<th>Nose</th>
<th>Eyes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>50.2</td>
<td>59.2</td>
<td>68.1</td>
</tr>
<tr>
<td>FTE</td>
<td>4.69</td>
<td>8.76</td>
<td>12.1</td>
</tr>
</tbody>
</table>

Table 1: Rank-1 recognition rate for the tested face regions. These results correspond to the best GMM configuration with \( M = 16 \).

3.3 Processing time

The proposed algorithm was developed in Python and tested on a PC with INTEL(R) Core(TM) i7-2600 processor and 8GB RAM. To train the GMM’s we used the scikit-learn toolbox, using diagonal covariance matrices for faster training. Finally the SIFT keypoint extraction was performed using the OpenCV 2.4.9 library. The computational time of a single identity test is dependent on the number of mixtures in the trained GMM’s. Considering \( M = 16 \), we observed a testing time of \( 0.0441 \pm 0.0073 \)s per identity. In our 100 individual subset this translates into a little below half a second to perform a full comparison with every individual enrolled in the model database. It might be interesting to check the real-time applicability of this method using online acquisition of face images.

4 Conclusion

In this work we presented a personal identification system based on the analysis of specific regions-of-interest – eyes, nose and mouth – which are detected using the Viola-Jones algorithm. The system performance has been examined in two ways: (1) how well does the Viola-Jones approach detect the ROIs, and (2) how accurate is the match between each image in the test dataset and the respective ID.

The obtained recognition results lead to a few main conclusions: 1) none of the explored sub-regions of the face was able of, per se, yielding results that match the state-of-the-art performance of full-face recognition. However, the eye region, known in literature as the periocular region proved to be, by a large margin, the most promising region for partial face recognition; 2) the results could have been better if we had not assumed that only the keypoint content of a scene is enough to describe it. By generating a likelihood based only on the presence of descriptors we are ignoring all the spatial information about their relative position.

Future work, besides exploring both the aforementioned topics, could focus on enabling live capture using a webcam, and how tracking mechanisms could be implemented to leverage the accuracy and processing time of our system. Another interesting question for further exploration is how the left and right eyes (which were treated as one single region in our approach) can be treated as independent models and how their synergy can help improve recognition.

Acknowledgment

This work is financed by the ERDF â€“ European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT â€“ Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281. The first author would also like to thank Fundação para a Ciência e Tecnologia (FCT) - Portugal the financial support for the PhD grant with reference SFRH/BD/87392/2012.

References

Mass detection on mammogram images: A first assessment of deep learning techniques

Inês Domingues
inesdomingues@gmail.com
Jaime S. Cardoso
jaime.cardoso@ieee.org

Abstract
Deep Learning approaches have gathered a lot of attention lately. In this work, we study their application to the breast cancer field, in particular for mass detection in mammograms. Several experiments were made on a real mammogram benchmark dataset. Deep Learning approaches were compared to other classification methodologies. It was concluded that, although useful, the implementation used does not outperform SVMs. Further study and adjustment of the method for this application is needed.

1 Introduction
Although the back-propagation algorithm [13] has been available for training neural networks for a long time, it was often considered too slow for practical use. As a result other learning models such as support vector machines (SVMs) dominated the field in the 1990s and 2000s. The term “deep learning” regained attention in the mid-2000s when it was shown that a many-layered neural network could be effectively pre-trained, one layer at a time, treating each layer in turn as an unsupervised restricted Boltzmann machine, and then using supervised back-propagation for fine-tuning [6].


In this work, we also present results on binary classification between cancer and non-cancer breast mass lesions extracted from the INBreast database, and go a step further by doing a preliminary study on the use of deep learning approaches to mass detection on mammogram images.

2 Methods
High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such “autoencoder” networks, but this only works well if the initial weights are close to a good solution.

Hinton and Salakhutdinov [7] describe an effective way of initializing the weights that allows autoencoder networks to learn low-dimensional codes that outperform principal components analysis as a tool for dimensionality reduction 1.

In this work we are interested in studying the behavior of the above mentioned method for the detection of breast masses in mammograms. The autoencoder consisted of an encoder with layers of size 1025-500-2000-2 and a symmetric decoder. The two units in the code layer were linear and all the other units were logistic.

3 Results
All 116 masses from the INBreast database [10] were used in the following tests. A rectangular ROI was generated from the bounding box (BB) of each mass, by expanding the BB by 20%. The examples with no masses were generated as follows. For each mammogram where a mass was extracted, an ROI of the same size was also randomly selected under the constraint that it did not intersect with the mass ROI. Every ROI was resized (using bi-cubic interpolation) to a square of 32 pixels per side. After resizing, pixel intensities were normalized to span the interval [0, 255]. While this approach avoids having to deal with the data unbalancing problem, it has the shortcoming that the selected non-mass patches may not represent every possible aspect of healthy breast tissue. In order to use all the available information, Curriculum learning approaches [2] where examples are not randomly presented but organized in a meaningful order might be studied in the future.

The dataset was split into training and testing in the proportion 75%/25%. In order to have more stable results, the split was repeated 40 times and the results were averaged. For comparison, the following methods were also used: k-Nearest neighbors (kNN); Decision Trees (DT); Linear Discriminant Analysis (LDA); Naive Bayes (NB); and Support Vector Machines (SVM). Matlab default values were used for every model parameter.

Throughout we speak of two results as being “significantly different” if the difference is statistically significant at the 1% level according to a paired two sided t-test, where each pair of data points consists of the estimates obtained in one of the 40 runs of the learning schemes being compared.

The first experiment concerns the distinction between mass and non-mass examples. In this experiment, only the above mentioned ROIs were used. Some patches with and without masses are shown in Figure 1 and quantitative results are presented in Table 1. Features learned by the first hidden layer of the deep learning method are depicted in Figure 2. All the

Figure 1: Examples of ROIs with (red) and without (green) masses.
Figure 2: Some filters learned by the first hidden layer of the deep learning method. Excitatory connections are shown in white, whereas inhibitory connections are in black.


differences have proved to be significant, except between Deep Learning and both kNN and NB. This means that the only method that is significantly better than Deep Learning is the SVM. Thus, in the remaining tests, besides Deep Learning, only SVMs will be tested.

In the next experiment, we try to adjust some parameters inherent to each model by using a two-fold cross validation methodology. For SVMs a grid search was performed over $C = 2^{-2 \ldots 2^{12}}$ and with three kernels, (1) Linear, (2) Gaussian Radial Basis Function (RBF) ($\gamma = 0.1 \ldots 1$) and (3) Polynomial ($\text{degree} = 2 \ldots 4$). Due to time constraints, for the Deep Learning experiments only the number of layers was varied from 1 to 3. Results can be seen in Table 2. For SVMs with the linear kernel, the results significantly improve over the ones presented in Table 1. Note that the default setting used was a Linear kernel with $C = 1$. Both the Polynomial and RBF behave significantly worse than the Linear kernel.

In the remaining experiments, when we refer to the SVM model, we mean using the Linear kernel and with grid search over the cost parameter $C$.

For Deep learning, it appears that increasing the number of layers has a positive effect. Differences between having 1 or 3 layers are significant. However, differences between having 1 or 2 layers or 2 or 3 layers are not significant. We will continue to use 3 layers in the remaining experiments.

Finally, we build up on the previous test and constructed a mass detection system. Each mammogram is scanned and the program finds the

Page 11 of 147
used to classify each patch as mass or non-mass. This scan is made in a multi-scale way by resizing the mammogram 20 times at different scales. For each tested patch, in case it is classified as a mass, a confidence value in the classification is also computed. As SVMs produce an uncalibrated value that is not a probability [11], the confidence value was set to 1. For deep learning, the continuous value of the output neuron is used as the confidence value. An SVM classifier was trained on these confidence values, in order to have a final binary classification per pixel. A formal analysis of the results is still in progress, but some selected detections can be seen in Figure 3 and.

![Some mass detection results using SVM classifier. Left: original mammogram with ground truth masses in green; Middle: SVM results with detected masses in red; Right: Deep Learning results with detected masses in red.](image)

Figure 3: Some mass detection results using SVM classifier. Left: original mammogram with ground truth masses in green; Middle: SVM results with detected masses in red; Right: Deep Learning results with detected masses in red.

It can be seen from the examples that both methods are able to find masses of different sizes and in different locations. It is important to note that no mass was missed by either technique. There are, however, some false positives. In order to decrease the quantity of false positives, some empirical rules can be implemented, these include eliminating very small regions, or regions whose height to width ratio is not reasonable. In alternative (or in complement) another layer of more sensitive classifiers could be built using only the detected regions.

4 Conclusions

In this work we have studied the problem of mass detection in mammograms. Several classifiers were tested and special attention was given to Deep learning methodologies. A strength of this work is that no “hand crafted” features were extracted. All methods worked directly in the patch pixels intensity space.

The encouraging results were obtained with no formal attempt of optimization of the hyper-parameters (e.g. the number of nodes per hidden layer). We believe that the selection of a different structure will further improve the results [1, 4].

Another technique that might improve the results would be to augment the dataset with known input deformations that are known not to change the class (e.g. small affine transformations such as translations, rotations, scaling, shearing) [3].

Some other possible applications of deep learning that we intend to study include: (1) detection of microcalcifications; (2) classification of suspicious lesions into benign/malign; and (3) to use the features learned by the autoencoder for the suspicious regions for Bi-RADS classification of the full mammogram image [5].

5 Acknowledgments

This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project 3D Models for Aesthetic Evaluation and Prediction of Breast Cancer Interventions with reference PTDC/SAU-ENB/114951/2009.

References


An Automatic Method for Assessing Retinal Vessel Width Changes

Behdad Dashtbozorg$^{1,2}$
behdad.dashtbozorg@fe.up.pt

Ana Maria Mendonça$^{1,2}$
amendon@fe.up.pt

Aurélio Campilho$^{1,2}$
campilho@fe.up.pt

$^1$INEB-Instituto de Engenharia Biomédica, Porto, Portugal
$^2$Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

Abstract

The Arteriolar-to-Venular Ratio (AVR) is commonly used in studies for the diagnosis of diseases such as diabetes, hypertension or cardio-vascular pathologies. This paper presents an automatic approach for the estimation of the Arteriolar-to-Venular Ratio (AVR) in retinal images. The proposed method includes vessel segmentation, vessel caliber estimation, optic disc detection, region of interest determination, artery/vein classification and AVR calculation. The method was assessed using the images of the INSPIRE-AVR database. A mean error of 0.05 was obtained when the method’s results were compared with reference AVR values provided with this dataset, thus demonstrating the adequacy of the proposed solution for AVR estimation.

1 Introduction

Retinal vessel features play an important role in the early diagnosis of several systemic diseases, namely diabetes, hypertension and vascular disorders. In diabetic retinopathy, the blood vessels often show abnormalities at early stages [8]. Changes in retinal blood vessels, such as significant dilatation and elongation of main arteries, veins, and their branches, are also frequently associated with hypertension and other cardio-vascular pathologies [2], [7].

Among several characteristic signs associated with retinal vascular changes, the Arteriolar-to-Venular Ratio (AVR) is used as an indicator of cardiovascular risk since it can reflect the narrowing of the retinal blood vessels. A low AVR value is associated with a high blood pressure, thus increasing the risk of stroke, diabetes and hypertension. Development of an automatic image analysis system for the estimation of AVR values requires vessel segmentation, accurate vessel caliber measurement, optic disc detection for region of interest delineation and artery/vein classification [3]. In this paper, we propose a fully automatic method for the estimation of the AVR value which achieves better performance than recently proposed approaches.

2 Material and methods

The estimation of AVR requires optic disc detection, vessel segmentation, accurate vessel caliber measurement and artery/vein classification [3]. Vessel segmentation must be used for finding the vessels, and optic disc detection is necessary to locate the region of interest (ROI) where vessel calibers are measured. An automatic AVR measurement system must classify the retinal vessels into arteries and veins with high accuracy since small classification errors can have a significant influence on AVR values. Finally, caliber measurements are used to compute AVR, according to the formula proposed by Knudtson et al. [3]. Figure 1 depicts the block diagram of the proposed method for AVR estimation. The main phases of the proposed method are described in the following subsections.

2.1 Vessel segmentation

For segmenting the retinal vessels, the method previously proposed by Mendonça et al. [5] was chosen and adapted for the segmentation of high resolution images [6]. Figure 2(b) illustrates the vascular tree for the original image of Figure 2(a).

Figure 2: (a) Original image; (b) Vessel segmentation result; (c) Region of interest for AVR (delimited by the two green circles) and the optic disc margin (red circle); (d) A/V classification result.

2.2 Vessel caliber measurement

For vessel caliber measurement a distance transform is applied to the segmented vascular tree and the result of this transform in each vessel pixel is the distance of the pixel to the closest boundary point, $d$. After that, for each vessel centerline pixel, the vessel caliber value, $vc$, is estimated by $vc = 2d - 1$.

Figure 1: Block diagram of the proposed method for AVR estimation.

2.3 Region of interest detection

The optic disc center (ODC) is located using an automatic methodology proposed by Mendonça et al. which is based on the entropy of vascular directions [4]; the region of interest (ROI) is centered on the ODC, and defined considering a disc diameter adapted to image resolution. Figure 2(c) shows an example of optic disc boundary and the region of interest for calculating the AVR.
2.4 Artery/vein classification

In order to classify the vessels as an artery or as a vein, an improved version of the automatic graph-based A/V classification method previously described in [1] is used. This method represents the segmented vasculature as a graph, which is afterwards modified for removing some typical errors and misrepresentations. Then the modified graph is analyzed for deciding about the type of intersection points (graph nodes) and based on the node types in each separate subgraph, all vessel segments (graph links) that belong to a particular vessel are identified and labeled using two distinct labels. Finally, intensity features are measured from the segments for assigning the final artery/vein class to each label. The result of A/V classification is shown in Figure 2(d).

2.5 AVR calculation

We have followed Knudtson’s revised formula [3] to calculate the AVR value. The Knudtson’s formulas for approximating the vessel equivalents are as the following:

\[
\text{Arterioles: } \hat{W}_a = 0.88 \times \left( w_{a1}^2 + w_{a2}^2 \right)^{0.5} (1)
\]

\[
\text{Venules: } \hat{W}_v = 0.95 \times \left( w_{v1}^2 + w_{v2}^2 \right)^{0.5} (2)
\]

where \( w_{a1}, w_{a2} \) and \( \hat{W}_a \) are, respectively, the widths of the narrower branch, the wider branch, and the estimate of parent trunk for arteries. \( w_{v1}, w_{v2} \) and \( \hat{W}_v \) have the same meanings for veins. For computing the central retinal artery equivalent, a set with the six largest arteries is first selected. Then, an algorithm is used for pairing up the largest and the smallest vessels in this set, and the parent trunk width is determined using Knudtson’s formula (1). A new set is formed by substituting the largest and smallest arteries with the parent trunk, and this algorithm is iterated until a single vessel is kept, whose width is the CRAE value. A similar procedure is used for calculating the CRVE, starting from the set formed by the six largest veins and using Knudtson’s formula (2). An approach similar to the one described in [9] is applied for AVR measurement. The ROI is equidistantly sampled to provide six regions. For each region, the six largest arteries and the six largest veins are identified, and the regional AVR value is obtained. The final AVR estimate for the complete image is the average of the six regional values.

<table>
<thead>
<tr>
<th>Table 1: Comparison of AVR values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Stdev</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>Max</td>
</tr>
</tbody>
</table>

3 Results

For validating the proposed AVR estimation method, we have used the 40 images of the INSPIRE-AVR dataset. This dataset includes two AVR measures that were computed by two ophthalmologists using a semi-automated computer program. The AVR estimates for Observer 1 are used as reference for calculating the errors for the results of both Observer 2 and our method. The analysis of the mean AVR and error values in Table 1 allows the conclusion that the error of our method is similar to the one of Observer 2 and is smaller than the error of the recent approach presented by Niemeijer et al. [9]. The agreement between the different methods and the reference can be observed in the Bland-Altman plots depicted in Figure 3. From the observation of these plots, it is worth mentioning that the results of our method are similar to those of Observer 2, and do not show a substantial bias as the mean of differences between the AVR values is close to 0. The 95% limits of agreement for all methods are also almost identical.

4 Conclusions

We have described an automatic method for calculating the AVR value in retinal images that is supported by a new solution for A/V vessel classification. The herein proposed approach was assessed in the images of the INSPIRE-AVR dataset where it has achieved a mean error of 0.05, identical to the one obtained by Observer 2. The low error is promising and demonstrates that our solution has a high potential for clinical application.

In order to determine the ROI for AVR calculation, a constant value is assumed for the optic disc diameter in all images. Therefore, as future work, we will focus on the development of a new approach for optic disc boundary detection and diameter measurement.

References


Abstract

In this work we developed a Computer-Aided Decision (CAD) system by making use of some physical properties in the image acquisition process and rebuilding the R, G and B channels achieving special photometric invariances. This is performed using a local descriptor such as the Scale Invariant Feature Transform (SIFT) and we assess the classifier performance by studying the impact of colour information to the descriptor. We achieved a performance of 79% with a regular gray conversion and 84% by making use of the opponent colours. The proposed set of descriptors achieved a sensitivity of 79% and a specificity of 89%.

Narrow-Band Imaging (NBI) is a promising tool in the diagnosis of cancer in gastroenterological examinations. It illuminates the mucosa with blue and green light and the fact that green penetrates deeper in the tissue will increase the contrast of superficial and deeper vessels and the visualization of certain structures that were unseen with conventional white light illumination. This brings new patterns that need to interpreted and analysed correctly to perform an accurate diagnostic. In this context a CAD Support System specialized for these images is crucial.

1 Introduction

NBI consists in narrowing the light output, illuminating the mucosa with blue and green with the use of special filters. Due to the different penetration depths of light (green penetrates deeper) and the fact that they match the absorption peaks of haemoglobin, blue will be absorbed by superficial vessels while green will be absorbed by deeper vessels thus enhancing the contrast between superficial and deeper vessels. Narrowing the bandwidth of the illumination also reduces scattering effects on the mucosal surface and thus the resulting image reveals structures and patterns that were unseen with a conventional white light illumination. Blue light is the input for the B and G channels and so the superficial vessels and structures will appear brownish while green is the input for the R channel and so deeper vessels will appear with a cyan colour.

These new patterns have a high correlation with histology as shown by Singh et al. They proposed a grading system for these images based on the vascular and structural pattern observed at the oesophageal mucosal surface:

- Type A: Round pits with regular microvasculature;
- Type B: Villous/ridge pits with regular microvasculature;
- Type C: Absent pits with regular microvasculature;
- Type D: Distorted pits with irregular microvasculature;

Figure 1 shows example images from our data set. We reduced the problem to a binary classification system by assigning Type A to class C1, the normal cases, and Types B, C and D to class C2, the cases that show an evolution to cancer. The main idea is to build a CAD system capable of separating normal cases from abnormal cases with the use of local descriptors.

2 Colour Invariant Features

2.1 Local Descriptor

In this work we propose to extract features using a local descriptor that shows invariance to scale, rotation and translation, the SIFT descriptor [2]. Patterns are spread over a wide region and for this reason the best way of sampling these images is to perform a dense sampling, as shown by Nowak et al. [4], placing a regular grid on top of the image and extracting the descriptors. The SIFT descriptor consists in computing the gradients in 8 possible orientations on an image patch of 4×4. The resulting vector is thus in a 128 dimensional feature space. The base descriptor is extracted with a regular conversion to gray-scale. In this work we intend to study the addition of colour information to this base descriptor and attend the possible improvements in the classification performance.

2.2 Adding Colour Information

The addition of colour to the base gray descriptor is performed by modulating the images according to the Dichromatic Reflection Model (DRM).

The Dichromatic Reflection Model The recorded image by a camera is the the sum of the incoming light at pixel position \( \rho \), \( L(\lambda, \rho) \), into the sensor, weighted by the spectral transmittance of the filter, \( \tau(\lambda) \) and the spectral responsivity of the camera, \( s(\lambda) \), over all wavelengths:

\[
C = \int_{\lambda} L(\lambda, \rho) \tau(\lambda)s(\lambda)d\lambda
\] (1)

This process of spectral integration is a linear transformation [5]. The DRM considers, for an inhomogeneous material and neutral interface reflection, that the formed image of an object can be seen as a combination of two terms: body and surface reflection, for more information see [5], and we can then consider the incoming beam as colour triple \( C = [R, G, B] \). The DRM states that:

\[
C = m_iC_i + m_bC_b
\] (2)

where the indices \( i \) and \( b \) denote interface and body respectively, \( m_i \) and \( m_b \) are the magnitudes of the corresponding reflections, \( C_i \) and \( C_b \) are the corresponding colours and are also colour triples.

Photometric Invariants To obtain photometric invariants we rebuilt the R, G and B channels according to the DRM. For simplicity we assume a matte surface, where the term \( m_i \) responsible for the specular reflections is negligible. In doing so we obtain the Lambertian model for diffuse body reflection:

\[
C = m_bC_b
\] (3)

A zero-order invariant can be obtained by normalizing the R, G, and B colours (rgb):

\[
r = \frac{R}{R+G+B} = \frac{m_bC_b^R}{m_b(C_b^R + C_b^G + C_b^B)} = \frac{C_b^R}{C_b^R + C_b^G + C_b^B}
\] (4)

Similar equations can be obtained for the normalized G and B channels. By doing so we are obtaining invariance to lighting geometry and viewpoint [1]. A set of first-order invariants can be obtained by taking the ratio of the channels (ratio). They have the same invariant properties as the normalized RGB colours.

\[
\begin{bmatrix} R \ R \ G \\ C_b^R \ C_b^G \ C_b^B \end{bmatrix}
\] (5)
In order to consider the quick density variations that cannot be captured linearly, we introduce the proportion under the logarithm (log) :

$$\left\{ \log \left( \frac{R}{G} \right), \log \left( \frac{G}{B} \right), \log \left( \frac{B}{R} \right), \ldots \right\}$$  \hspace{1cm} (6)

Human perception of colours relies on the opponent process theory. For this matter we consider the opponent colour space (opp):

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{R+G+2B} \\ \frac{R+G-2B}{R+G+2B} \\ \sqrt{\frac{2}{3}} \end{pmatrix}$$

The $O_1$ and $O_2$ channels are independent to highlights but are still sensitive to geometry, shading and the intensity of the light source. The $O_3$ channel is the intensity and contains no invariance properties. We can also take the hue for the opponent colours (hue):

$$hue = \arctan \left( \frac{O_1}{O_3} \right) = \arctan \left( \frac{\sqrt{3}(C_b^G-C_b^B)}{C_b^R+C_b^G+C_b^B} \right)$$  \hspace{1cm} (8)

The chromatic opponent colours (chroma) can also be computed.

$$\begin{pmatrix} C_a \\ C_b \end{pmatrix} = \begin{pmatrix} \frac{O_1}{O_3} \\ \frac{O_2}{O_3} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{2}{3}} (C_a^G-C_a^B) \\ \sqrt{\frac{2}{3}} (C_a^R-C_a^B) \end{pmatrix}$$  \hspace{1cm} (9)

Both the hue and the chromatic opponent colours are invariant to lighting geometry and specularities [1].

3 Image Representation and Pathology Recognition

For the recognition of pathologies through local information, k-means method shows to be the most robust for the image description. This method is also known as the Bag of Words. As a brief description lets suppose we have a set of data points $\{x_k\} \in \mathbb{R}^d$, where $d$ is the dimension of the feature space. K-means consist in searching for $K$ cluster centres, the vectors $\{c_k\} \in \mathbb{R}^d$ and finds an assignment function $A : \{x_k\} \rightarrow \{c_k\}$ until the sum of the squares of the distances to each data point to its closest vector $c_k$ is minimum. We estimated an optimal value of 400 clusters. For bigger values the computational cost is not justified.

Learning the Patterns This is performed with a Support Vector Machine (SVM). They rely on the Support Vectors to determine a maximum margin hyperplane that separates the classes. Let us assume we have $N$ input vectors $x_1, x_2, ..., x_N$ with the corresponding labels $y_1, y_2, ..., y_N$ where $y_n \in \{-1, 1\}$. The data points are transformed into a higher dimensional space by means of a linear mapping, $\phi(x)$. The maximum-margin hyperplane is defined by $g(x) = w^T \phi(x) + b$, where $w$ is the weight vector and $b$ is a bias parameter. To find the maximum margin is equivalent to find:

$$\min \frac{1}{2} ||w||^2 + C \sum_{n=1}^{N} \xi_n$$  \hspace{1cm} (10)

subject to the constraints:

$$y_n(w^T \phi(x_n) + b) - 1 + \xi_n \geq 0 \hspace{0.5cm} \forall \hspace{0.5cm} n \in \{1, ..., N\}$$  \hspace{1cm} (11)

where the parameter $C$ controls the trade-off between the slack variable penalty ($\xi_n$) and the size of the margin. In this work we estimated that, between the Intersection, the $\chi^2$ and the Linear Kernel, the best results were attained with the Intersection Kernel.

4 Results

The image set is composed of endoscopic images from the Barrett’s oesophagus with a total of 250 images classified according to Singh’s proposal and they were also annotated. From those, 61 are from Type A (normal), and the remaining 189 are from Types B, C and D (pathologic), reducing the problem to a binary classification system. For the training process we considered 50 normal and 50 pathologic images in order to balance the dataset. To attain the best regularization parameter, we performed a three-fold cross-validation scheme. We began by analysing the performance with the gray descriptor. Each step consist in adding consecutively colour information to this descriptor according to Section 2.2. In order to obtain stable results we repeated each experiment 10 times and calculated the mean values of the Mean Average Precision (MAP) and the corresponding standard deviations.

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>{gray}</td>
<td>0.794 ± 0.038</td>
</tr>
<tr>
<td>{opp}</td>
<td>0.748 ± 0.083</td>
</tr>
<tr>
<td>{gray, rgb}</td>
<td>0.819 ± 0.069</td>
</tr>
<tr>
<td>{gray, rgb: ratio}</td>
<td>0.815 ± 0.050</td>
</tr>
<tr>
<td>{gray, rgb: ratio, log}</td>
<td>0.805 ± 0.049</td>
</tr>
<tr>
<td>{gray, rgb: ratio, log, opp, chroma, hue}</td>
<td>0.840 ± 0.042</td>
</tr>
</tbody>
</table>

Table 1: MAP for the tested sets of descriptors.

From Table 1 we observe that the set $\{gray, rgb: ratio, log, opp\}$ gives the highest MAP. For the referred set of descriptors we present the sensitivity and specificity in Table 2.

<table>
<thead>
<tr>
<th>Descriptor set</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>${gray, rgb: ratio, log, opp}$</td>
<td>0.788</td>
<td>0.891</td>
</tr>
</tbody>
</table>

Table 2: Sensitivity and Specificity for the best set of descriptors.

5 Conclusions

The obtained results show that adding information from the RGB channels independently, the ratio of the RGB channels, the logarithmic of these ratios and the information from the opponent colours gives the best results. With this set of local descriptors we attained a MAP of 84.0%, a sensitivity of 78.8% and a specificity of 89.1%. These high values obtained allow to conclude that the use of local descriptors such as SIFT actually give very good results mainly due to its invariance properties and to the very special conditions that these images are acquired. The addition of colour information improved the classifier performance in about 5% comparing to the gray descriptor. The fact that the opponent colour system highly decorrelates the RGB channels explains this gain in performance.

Acknowledgments

This work was financially supported by FCT (Portuguese Science Foundation) grant PTDC/EIA-CCO/109982/2009. We also would like to thank project PEst-OE/EEI/LA0008/2013.

References

Abstract

With the recently emergence of the high-throughput sequencing technologies, full genomic DNA sequences have been released to the public, allowing for the first time large genomic computational studies between species. One of those studies is chromosomal distances between Hominidae primates in order to unveil insights of evolution.

The most successful and promising distance metrics seems to rely on the compressed based type. Therefore, we propose and use an admissible normalized compression distance, based on a specific conditional compression analogue which internally is composed by a mixture of static and dynamic finite-context models.

1 Introduction

The recent full genomic sequencing of primates (Hominidae) have brought new challenges, such as the identification of chromosomal distances in order to find insights of large genomic evolution. Although the existence of several distance metrics, such as Hamming and Levenshtein, the compressed-based distances seem to be the most successful ones.

A compression-based distance measure assesses the distance between two objects using the number of bits needed to describe one of them when a description of the other is available. The foundations of compression distances are built upon the Kolomogorov notion of complexity, also known as algorithmic entropy, where expression distances are formalized as

$$H(x) = \text{length of shortest binary program for } x\$$

and

$$C(x) = \text{length of shortest binary program for the reference prefix Turing machine that, on input } x \text{ outputs } x\$$

and

$$\lambda = \text{length of shortest binary program for } \lambda\$$

where, asymptotically, $C(x) = C(x) - \lambda$, where $C(x) \leq C(x) + C(\lambda) = C(x) + C(y) = C(x) + C(x)$. For stationary sources, we could compute weights such that $\sum_i w_i = 1\$.

2 Method

2.1 Admissible compression distance

Instead of using the conventional conjoint compression to deduce the mutual information content in (2), we have created a specific analogue of a conditional compressor, $C(x|y)$ (see next subsection), in order to compute (1). Therefore, the admissible NCD can be calculated using

$$\text{NCD}(x, y) = \frac{\max[C(x|y), C(y|x)]}{\max[C(x), C(y)]},$$

(3)

where, asymptotically, $C(x|y) = C(x), C(x) = 0, C(y) \leq C(x), C(x) + C(y) = C(x) + C(x)$ and $C(x) + C(y) \geq C(x, z) + C(y, z)$, up to an aditive logarithmic term.

2.2 Conditional compressor

We have conducted a NCD compressor based on a mixture of two classes (static and dynamic) with multiple finite-context models. Accordingly, the compression is performed in two phases. In the first phase, the static class of finite-context models, with variable orders, accumulate the counts regarding the $y$ object. When the entire $y$ object is processed, the models are kept frozen and, hence, the second phase starts. At this point, the $x$ object starts to be compressed using the static models, from the first phase, in cooperation with the new multiple finite-context models, of variable orders, that dynamically accumulate the counts only from the $x$ object.

The per symbol information content average provided by the finite-context model of order-$k$, after having processed $n$ symbols, is given by

$$H_k = -\frac{1}{n} \sum_{i=0}^{n-1} \log_2 P(x_{i+1}|x_{i-k+1}).$$

(4)

where “bpb” stands for bit per base. The process of supervision is held by mixture weights which relate each static and dynamic model. Therefore, the probability estimate can be given by a weighted average of the probabilities provided by each model, according to

$$P(x_{n+1}) = \sum_k P(x_{n+1}|x_{n-k+1}) w_{k,n},$$

(5)

where $w_{k,n}$ denotes the weight assigned to model $k$ and $\sum_k w_{k,n} = 1$.

For stationary sources, we could compute weights such that $w_{k,n} = P(k|x_1), i.e., according to the probability that model $k$ has generated the sequence until that point. In that case, we would get

$$w_{k,n} = P(k|x_1) = P(x_1), k),$$

(6)

where $P(x_1)$ denotes the likelihood of sequence $x_1$ being generated by model $k$ and $P(k)$ denotes the prior probability of model $k$. Assuming $P(k) = \frac{1}{K}$, where $K$ denotes the number of models, we obtain $w_{k,n} = P(x_1, k)$.

Calculating the logarithm we get

$$\log_2 P(x_{i+1}|k) = \log_2 \prod_{i=1}^{n} P(x_i|x_{i-1}) = \sum_{i=1}^{n} \log_2 P(x_i|x_{i-1}).$$

(7)

which corresponds to the code length that would be required by model $k$ for representing the sequence $x_{1:n}$. It is, therefore, the accumulated measure of the performance of model $k$ until instant $n$. However, since the DNA sequences are not stationary, a good performance of a model in a certain region of the sequence might not be attained in other regions. Hence, the performance of the models have to be measured in the recent past of the sequence, for example over a window of appropriate size, or be equipped with a mechanism of progressive forgetting of past measures. We opted for the latter possibility, using the recursive relation

$$\log_2 P(x_1, k) = \log_2 \prod_{i=1}^{n} P(x_i|x_{i-1}) = \sum_{i=1}^{n} \log_2 P(x_i|x_{i-1}).$$

(8a)

As can be verified, this relation corresponds to a first-order recursive filter that, for $\gamma \in [0, 1)$, has a low-pass characteristic and an exponentially decaying impulse response. For more information on finite-context modelling and mixtures see [8, 9].
Table 1: Data set table. The number of expected chromosome pairs for each species is represented by 'Exp', while 'Missing' is a nonexistence sequence and Mb represents the approximated size in Mega bases.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Build</th>
<th>Exp</th>
<th>Missing</th>
<th>Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>37:p10</td>
<td>23</td>
<td>-</td>
<td>2,3861</td>
</tr>
<tr>
<td>Pan troglodytes</td>
<td>2.14</td>
<td>24</td>
<td>-</td>
<td>2,756</td>
</tr>
<tr>
<td>Gorilla gorilla</td>
<td>1.00</td>
<td>24</td>
<td>Y</td>
<td>2,719</td>
</tr>
<tr>
<td>Pongo abelii</td>
<td>1.3</td>
<td>24</td>
<td>Y</td>
<td>3,028</td>
</tr>
</tbody>
</table>

Figure 1: *P. troglodytes*, *G. gorilla* and *P. abelii* inter-genomics chromosomal NCD heatmaps in relation to *H. sapiens*.

3 Results

The data set is composed by 4 genomes (Table 1), downloaded from the NCBI (ftp://ftp.ncbi.nlm.nih.gov/Genomes).

In Fig. 1 the inter-chromosomal NCD distance heatmaps for the three species relatively to *H. sapiens* have been plotted in an approach all with all. As it can be seen, for all species there is a direct correlation with the respective chromosomal number, with the exception of chromosome 2 (related to 2A and 2B). This is justified by a presumed chromosomal fusion in humans from previous ancestors [5].

Moreover, the human Y chromosome is highly related with the X chromosome of all addressed species, namely for the *P. troglodytes*, because the Y chromosome exchanged genetic information with X in the recombination process [3]. Furthermore, there is a low distance between chromosomes 5 and 17 of *G. gorilla* and *H. sapiens*, justified by a chromosomal translocation [10].

On the other hand, in Fig. 2 are presented the chromosomal distances of *P. troglodytes*, *G. gorilla* and *P. abelii* (chromosomes 2A and 2B have been concatenated) according to *H. sapiens* chromosomes order. At glance, *P. troglodytes* has got the lowest distance relatively to *H. sapiens*, and after *G. gorilla* and *P. abelii*, respectively. Specifically, *G. gorilla* chromosomes 5 and 17 have large distances because of the previous mentioned translocation, while *P. abelii* seems to have a very different chromosome 1, besides other relevant dissimilarities.

According to [4], besides the high divergence of Y chromosome, there are several breakpoints in chromosomes 4, 5 and 12, which were tested by fluorescence in situ hybridization (FISH), in *P. troglodytes* using *H. sapiens* as reference. Fig. 2 reports the same dissimilarities, surprisingly adding chromosome 17.

Finally, we have found that chromosomes 4, 12 and 18 of *G. gorilla* have lower distances to *H. sapiens* than to the respective *P. troglodytes* chromosomes, while chromosomes 5 and 17 of *G. gorilla* have higher distances relative to that of *P. abelii*. Mitochondrial sequences, as expected, show that *P. troglodytes* is the nearest *H. sapiens* species, followed by the *G. gorilla* and, lastly, by *P. abelii*.

4 Conclusions

An admissible normalized compression distance has been proposed, based on a specific conditional compression analogue. The compressor is constituted by a set of multiple static and dynamic finite-context models that are supervised by a mixture model.

We have addressed a study on chromosomal distances between Hominoidea primates, agreeing with several already documented results (using other approaches), but also unveiling undocumented ones.

The biggest advantage of this method is the process automation for any kinds of genomic DNA sequences, while using biological techniques can be a rather lengthy process.

Acknowledgements

This work was supported in part by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Foundation for Science and Technology, in the context of the projects FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011) and Incentivo/EEI/UI0127/2013.

References

Comparative study of two movement identification strategies on Brain-Computer Interface motor task

Mariana Branco
mariana.branco@ist.utl.pt
João Sanches
jmrs@ist.utl.pt
Rodrigo Ventura
rodrigo.ventura@isr.ist.utl.pt

Institute for Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal

Abstract

Support Vector Machine (SVM) is a state-of-art machine learning algorithm broadly used classification and learning tasks. It has been recently used on Brain-Computer Interface (BCI) systems to discrimination between motor tasks of electroencephalography (EEG) signals. However, there are different possible strategies to implement the SVM classifier. In this paper it is compared two SVM strategies to classify a motor movement task: 3-class and binary, hierarchical SVM. One healthy subject was submitted to an offline BCI experiment and the results show that the binary classifier reports in average higher classification accuracies than the 3-class SVM (68.47% against 53.07%). The difference between both strategies is statistically significant (p-value < 0.001; for 0.05 significance level).

1 Introduction

Support Vector Machine (SVM) is a linear classifier first developed by Vapnik in 1995 [7]. The application of the SVM method has only recently begun, however some state-of-art performances have already been reported [3]. In particular, for Brain-Computer Interface the SVM is of particular interest for motor tasks classification, e.g. [2], [6] and [9]. This algorithm has the advantage of achieving very good classification accuracy and easy implementation of non-linear classification functions [4], [5]. Nonetheless, the feature classification requires proper design, since excessive adaptation of the algorithm can impair performance [8]. Hence, this paper compares two SVM strategies for motor imagery task classification: 3-class and binary SVM.

1.1 SVM formulation

Support Vector Machines (SVMs) is implemented here using the LibSVM library available online [1]. The LIBSVM library is used to obtain a model after training a data set and then use it to predict information of a testing data set. There are several types of formulations available, however here it is only used the 𝜈-support vector classification (𝜈-SVC).

The SVM classifier is based on the class of hyperplanes

\[ (\mathbf{w} \cdot \mathbf{x}) + b = 0, \mathbf{w} \in \mathbb{R}^N, b \in \mathbb{R} \]  (1)

where \( \mathbf{x} \) is a N-dimensional patterns with class label \( y \). These hyperplanes correspond to the decision functions

\[ f(x) = \text{sign}((\mathbf{w} \cdot \mathbf{x}) + b) \]  (2)

In general, it maps the input space nonlinearly into a higher dimensional feature space and constructs a separating hyperplane with maximum margin there. By solving a constrained quadratic optimization problem with \( w = \sum v_i x_i \) as solution, it can be uniquely constructed the optimal hyperplane (one with maximal margin of separation between two classes). The training patterns are therefore called support vectors and carry all information about the classification problem. For more details see [3].

2 Methods

2.1 Study Design

A scalp EEG was recorded from a female healthy volunteer (21 years old). The experiment last for one hour including the preparation of the EEG setup and cap collocation. The subject was seated comfortably in from of a computer screen (1 m distance) and was asked to reduce movements to minimal during the sessions and concentrate on the screen.

It was recorded one session of 20 cued trials of ‘left’ hand and ‘right’ hand movement. Thus, it were defined 3 classes: 1) ‘idle’; 2) ‘right’ hand; and 3) ‘left’ hand. Each trial started with a fixation cross, followed by a ‘beep’ to alert the subject for the ensuing cue. At 3 s a letter would appear on the screen, either ‘L’ for left hand movement or ‘R’ for right hand movement (see Figure 1), lasting for 5 s. The EEG was recorded continuously through out the 20 trials. At 8 s the subject would have a random inter-trial interval of minimum 0.5 s and maximum 2 s. During this period the subject could blink and move freely. The fixation cross was continually displayed on the screen during the entire session. Figure 2 depicts the timing scheme of the experimental paradigm.

Figure 1: The fixation cross a) was displayed in the beginning and end of each trial. At 3 s a letter would appear either on the left, b), or right sided, c), of the screen to indicate which hand to move.

Figure 2: Experimental paradigm for the offline analysis. The EEG was continuously recorded during the session, however only the period marked as ‘acquisition’ was processed.

The data was recorded with a 64-channel BioSemi© ActiveTwo system, composed by a AD-box, a battery box, a USB2 receiver and the ActiView© software. The electrode cap displays the electrodes on the 10-20 international placement system. The EEG signals were acquired with a sampling frequency of \( f_s = 512 \text{ Hz} \) and downsampled to \( f_s = 128 \text{ Hz} \). From the 64 channels, 14 were removed due to severe noise or higher susceptibility to artefacts. The trials were defined using a FieldTrip function as well as the data preprocessing: a 40th order Butterworth bandpass filter with frequency range \([f_{low}, f_{high}] = [3, 50] \text{ Hz} \); and channel referencing with Common Average reference (CAR).

2.2 Feature extraction

Power features are extracted by squaring the discrete Fourier coefficients (Parseval’s theorems). The fast fourier transform (FFT) coefficients were extracted, squared and averaged in the mu band (7-14 Hz):

\[ \text{Power}_{\text{mu}}(f) = \left( \text{FFT}(\text{EEG}(ch, n)) \right)^2 \]  (3)

where \( ch \) an EEG channel and \( n \) is a time sample index.

2.3 Classification implementation

To obtain the predicted label it was implemented manually the algorithm explain in 1.1. This function replicates the results obtained by the libSVM
dict function available in the LIBSVM library, however it is less computationally expensive. The svmModel structure is saved after training the model with the LIBSVM training function.

Two SVM strategies were compared: 3-class SVM and binary, hierarchical SVM. The former classified the features according to the labels 0 (‘idle’), 1 (‘right’ hand) and 2 (‘left’ hand). The second strategy involves two binary classifiers as represented in Figure 3. The first classifier (C1) categorizes the data as ‘moving’ or ‘idle’ (or ‘not move’), whereas the second classifier (C2) classifies between ‘left’ or ‘right’ hand. A 10-fold cross-validation (CR) was used for training and testing of each model.

Figure 3: Hierarchical SVM classifier for all classes. Each orange dot corresponds to a classifier. The first level (C1) classifies between ‘Move’ or ‘Not Move’ (or ‘Idle’) and the second level (C2) classifies between ‘Left’ or ‘Right’ hand.

3 Results

The outcome accuracies from the 10-fold CR are depicted in Figure 4 and 5 for 3-class and binary SVM, respectively. For 3-class SVM the average classification was 53.07% (standard deviation σ = 4.71%), whereas for binary SVM was 76.56% for C1 and 60.37% for C2 (σ1 = 1.09% and σ2 = 3.37%).

![3-Class SVM classifier](image)

Figure 4: Classification accuracies for a 10-fold CR of 3-class SVM.

![Binary and hierarchical SVM classifier](image)

Figure 5: Classification accuracies for a 10-fold CR of C1 (dark blue) and C2 (light blue) classifiers. The orange line represents the average of both classifiers for each CR repetition.

A paired t-test was used to test the statistical null hypothesis: the 3-class SVM and binary SVM come from a normal distribution with equal mean and unknown variance. Comparing C1, C2 and the average of both with equality SVM, a p-value < 0.001 (for 0.05 significance level) was obtained, inferring that binary SVM reports statistically higher accuracies than the 3-class SVM.

4 Discussion

When designing a SVM classifier it is necessary to take into consideration its formulation. Although it is theoretically a binary classifier, its implementation in [1] can accept n classes (n > 2). When classifying motor tasks power features, the results show that in average the binary, hierarchical classifier reports higher accuracies that the 3-class SVM. In average, the binary classifier reports 76.56% (C1) and 60.37% (C2) accuracies. Both C1, C2 and the average of both (68.47%) are statistically higher than the 3-class SVM (average 53.07%; p-value < 0.001, for 0.05 significance level). Comparing Figure 4 and Figure 5, it is visible that the classification accuracies for C2 are considerably lower than C1. Thought, both binary classifiers reveal to be significantly higher than the 3-class SVM.

Besides the highest accuracies, the binary classifier has also the advantage of separating the motor task into two sub-tasks: ‘move’ or ‘not move’; and if ‘move’, ‘left’ or ‘right’ hand (see Figure 3). This particularity is especially useful for BCI analysis since it is possible to discriminate how each classifier (C1 and C2) behaves. Moreover, it allows the selection of which features to be classified by C2, decreasing the error of classifying ‘move’ labels that are in fact ‘idle’.

5 Conclusion

In this paper two SVM strategies were applied to a motor movement offline BCI experiment. Power features were extracted and classified using two SVM strategies. The binary, hierarchical classifier reports higher accuracies when compared to 3-class SVM. This result is of extremely relevance to design BCI systems, improving the accuracy results decrease the accuracy error and speeding the system processing.

References


RecPad 2013

EEG time-frequency analysis for ERD/ERS temporal pattern characterization on brain computer interface motor task

Mariana Branco
mariana.branco@ist.utl.pt

João Sanches
jmr@ist.utl.pt

Fernando Lopes da Silva
F.H.LopesdaSilva@uva.nl

Institute for Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal

Abstract

One of the most important aspect in designing a Brain-Computer Interface (BCI) system is the feature extraction. Event-Related (De)Synchronization (ERD/ERS) phenomena are widely used features in this respect. However, before extracting the features it is necessary to verify if the data present well defined ERD/ERS patterns. In this paper we used a time-frequency (TF) analysis to recognize and identify event-related patterns. One healthy, female subject was submitted to a BCI offline experiment. The results of the TF analysis of the data reveals the presence of ERD/ERS patterns as reported in the literature.

1 Introduction

In general a non-invasive BCI system consists essentially of five elements (e.g. [1], [7], [8] and [9]). Succinctly, it is acquired using a non-invasive imaging technique, like electroencephalography (EEG), and the most valuable features to be applied, are extracted.

Event-related Potentials (ERPs) are time-locked potentials that change after a particular event whether external or internal: a sensory stimulus, a mental event, or even the omission of a stimulus within a series of repetitive stimuli. In many cases of BCI applications changes in ongoing EEG signals are used. On the one hand, a decrease of EEG power in a given frequency band is assumed to result from a decrease of synchrony of neurons and is called Event-related Desynchronization (ERD). This is commonly seen in the alpha frequency band corresponding to a state where the corresponding cortical areas are actively involved in attentional processes and/or motor preparation. On the other hand, an increase of power in a given frequency band, that is assumed to reflect an increase in neuronal synchrony, corresponds to a state where inhibition of the cortical network prevails, and is called Event-Related Synchronization (ERS). These ERD/ERS phenomena in the alpha oscillations are typically recorded over the somatosensory cortical areas (mu rhythm) and have to be well defined. In this paper we used a time-frequency (TF) analysis (see [2]) as a tool to identify event-related EEG patterns in an offline motor movement based BCI experiment.

2 Method

A scalp EEG was recorded from a healthy female subject (21 years old). The experiment lasted for about an hour including the placement of the EEG cap, that displays the electrodes on the 10-20 international placement system. The subject was seated comfortably in front of a computer screen and was asked to refrain from making movements during the sessions and to concentrate on the computer screen. One session comprised of 40 cued trials was recorded. In 20 trials the subject was asked to imagine a left hand movement, and in other 20 to imagine a right hand movement. Each trial started by asking the subject to fixate a cross in the computer screen, that was continuously displayed on the screen during the entire session. Thereafter, a ‘beep’ sound was presented to warn the subject for the ensuing cue, that appeared 3 s later. This could be either the letter “L”, for left hand movement, or the letter “R” for right hand movement (see Figure 1); the letter presentation lasted 5 s. At 8 s the subject was allowed to blink and move freely during an inter-trial period that lasted between 0.5 s and 2.0 s. Figure 2 depicts the timing scheme of the experimental paradigm.

During the whole period the EEG was continuously recorded with a 64-channel BioSemi® ActiveTwo system, composed by a AD-box, a battery box, a USB2 receiver and the ActiView® software. EEG signals were acquired with a sampling frequency of $f_s = 512$ Hz and downsampled to $f_{s} = 128$ Hz.

The interface built in Matlab® was synchronized with the acquisition system through markers automatically positioned on the ActiView® software (parallel port communication). The last seconds of each trial were not considered for feature extraction, removing the eye blink and EMG artefacts. The period of 3 s before the cue constituted the baseline of the respective trial. The trials were defined using a FieldTrip function as well as the data preprocessing: a 4th order Butterworth bandpass filter with frequency range $[f_{\text{low}}, f_{\text{high}}] = [7, 14]$ Hz; baseline correction by subtracting the mean of the whole trial; and channel referencing with Common Average reference (CAR).

2.1 Time-Frequency Analysis

Time-frequency (TF) analysis was performed for the whole trial, from 0 s to 9 s, i.e. including the baseline and cued periods, applying a baseline correction based on the window period from 0-3 s. The frequency range analysed was from 7-14 Hz (mu rhythm), using 0.5 s time-windows sliding from 0 s to 9 s in steps of 0.05 s (90% overlap). The power estimation is averaged across trials (20 per motor task) and is given in percentage. According to the literature the channels C3 and C4 overlying the cortical sensorimotor hand area provide the most important signals for hand motor movement and imagery [3].

$$\text{Power}\% = \frac{\text{Power} - \text{Power}_{\text{Baseline}}}{\text{Power}_{\text{Baseline}}} \times 100 \quad (1)$$

In order to see the evolution of the power in time and the averaged power during the baseline, two figures (one for each hand) were added. Each figure contains three plots with the averaged power (across frequency and trials), the difference between power in channel C3 and C4, and the standard deviation of the averaged power. All three plots are in function of time (0 − 9 s).

3 Results

Left Hand: Figures 3 and 4 show a desynchronization in the channels C3 and C4, specially for the lower frequencies (11-14 Hz). Figure 5 show the averaged power (‘AVG’; normalized by the mean power of the baseline period), in the time window 0-9 s, as well as its standard deviation.

- Figure 1: The fixation cross a) was displayed in the beginning and end of each trial. At 3 s a letter would appear either on the left, b), or on the right sided, c), of the screen to indicate which hand to imagine or move.

- Figure 2: Experimental paradigm for the offline analysis. The EEG was continuously recorded during each session, however only the period marked as ‘acquisition’ was processed.

---

Footnote: The work was supported by the FCT project [PEst-OE/EEI/LA0009/2013].
(‘ST’). Looking to the upper and lower frequencies, it is not possible to clearly identify the ERD on channel C4 and ERS on channel C3, being evident an ipsilateral desynchronization. In Figure 5, the high average and standard deviation values during the baseline is noticeable, followed by a significant decrease after the cue is presented (4 s).

In the upper panels, the horizontal line marks the level of baseline power and the vertical line the cue onset.

**Figure 5: Time-Power plot of channel C3 and C4, right hand. ‘AVG’ is the averaged power and ‘ST’ its standard deviation (N=20). Lower frequencies range from 7-10 Hz, while upper frequencies from 10-14 Hz.**

**Right hand:** Comparing the channels C3 (Figure 6) and C4 (Figure 7) a higher desynchronization on C3 than on C4 is observable, apparent for both the upper and lower frequencies of the mu rhythm. Considering the averaged power and its standard deviation (Figure 8), for lower frequencies the average power in channel C3 is considerably lower than in channel C4, though the same pattern can also be observed for upper frequencies. Despite the high variability during the baseline, the power stabilized after the cue onset, as well as its standard deviation.

**Figure 6: Time-frequency plot of channel C3, right hand (N=20).**

**Figure 7: Time-frequency plot of channel C4, right hand (N=20).**

**5 Conclusion**

In order to verify if the data contains ERD/ERS patterns, a TF analysis was performed. This study explains how to use TF information of upper and lower frequency components of the mu rhythm to identify the ERD/ERS phenomena. For hand motor tasks the power over channels C3 and C4 are analysed and the results seem to agree with the ones found in the literature. Hence, the data can be consider for further analyses, namely to apply feature extraction methods of ERD/ERS.

**References**


Abstract

In this work we study the impact of a set of bag-of-features strategies for the recognition of cancer in gastroenterology images. By using the Scale Invariant Feature Transform (SIFT) descriptor, we analyzed the importance and performance impact of term weighting functions for the construction of visual vocabularies. Further analyzes were conducted in order to ascertain the robustness of multiclass decomposition rules for Support Vector Machines with different kernels. Our study was extended by tailoring a decomposition rule that explores prior knowledge according the four grades of the Singh taxonomy (SDR). We found that SDR coupled with a frequency term weight function attained the best overall results (80%) when trained with an intersection kernel.

1 Introduction

Although clinical literature shows that early detection via endoscopy typically leads to a good prognosis, preventing the cancer from evolving into more dangerous stages, its detection is still difficult. This is mostly due to the visual cues that cancer exhibits in its early stages [3]. In this work we analyze the benefit of the Scale Invariant Feature Transform (SIFT) for recognition of cancer on images from the Barrett’s esophagus. Moreover, we also consider the relevance of term weighting approaches in this context regarding the different multiclass SVM decomposition strategies. Finally, in this paper it is also proposed a decomposition strategy by incorporating prior knowledge from our problem. This strategy also targets the training computation complexity reduction while at the same time reducing dubious decisions generated by standard decomposition rules.

2 Related Works

Local image analysis techniques have shown their importance for the detection of interest points that can be identified under different illuminant and shape conditions. In this Section we will briefly detail the established framework for local image analysis.

Local Descriptor: Scale Invariant Feature Transform (SIFT) is a well-known technique for feature description with a different number of applications already reported in the literature. After applying an interest point detector, such as a LoG (Laplacian of Gaussian) or a DoG (Difference of Gaussian), for instance, the SIFT descriptor is commonly built upon the measurements of the gradient in 8 different orientations in an image patch of size 4 x 4. As a result, a feature vector of size 128 is obtained.

Visual Words: For the visual words construction, a K-means technique is mentioned to be the preferable approach. In a nutshell, K-means is a simple unsupervised approach that cluster data into K groups. For sake of simplicity and since it is not the focus of this work, next we briefly describe this step. Given an initial arbitrary location for the clusters, a K-means algorithm iteratively finds new centers locations covering as best as possible data dispersion until a stopping criterion is achieved. The right choice for the vocabulary size does not follow a strict rule being a trade-off between generalization and discriminant capability [4].

Term weight functions: Borrowed from information retrieval literature [2], term weight functions have been widely used in many different computer vision studies [4]. This process follows the visual words construction step (described in the previous paragraph) in order to build the visual vocabulary. This is performed by counting the number of descriptors associated to a visual word. In [4] an extensive evaluation with different term weight functions was conducted. Although documented comparisons usually deal with term frequency (txx, number of times visual word $f_i$ is present) and other variants, in the context of gastroenterology images such analysis has been disregarded. Here we study the impact of the binary term weight function (txx, 1 if visual word $f_i$ is present, 0 otherwise) and term frequency weight function, txx.1

Learning: Support Vector Machine (SVM) is a popular learning mechanism. In its simplest form, SVM uses a linear separating hyperplane to create a binary classifier with a maximal margin. In cases where data are not linearly separable, they are transformed to a higher dimension than the original feature space. Such is done by choosing a given kernel function, representing the inner product in a implicit higher dimension space. More formally, given the training set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ with input data $x_i$ and corresponding binary class labels $y_i \in \{-1, 1\}$, the maximum-margin hyperplane is defined by $\langle g(x) = w^T \varphi(x) + b \rangle$ (where $\varphi(x)$ denotes a fixed-feature space transformation and $b$ a bias parameter). $x_i$ is assigned to class 1 if $g(x_i) > 0$ or to −1 if $g(x_i) < 0$. The maximization of the margin is equivalent to solving:

$$\min_{w,b} \frac{1}{2}w^Tw + C\sum_{i=1}^{N} \xi_i \quad \text{s.t.} \quad y_i\langle w^T \varphi(x_i) + b \rangle \geq 1 - \xi_i, i = 1, \ldots, N$$

The slack variables $\xi_i, i = 1, \ldots, N$ are introduced to penalize incorrectly classified data points. Due to the histogram nature of features used in this framework, researchers have identified that histogram kernels are more effective than the common ones used in SVMs. Our study was extended to analyze the impact of the kernel selection for recognition of cancer using the common linear kernel $K(x_i, x_j) = \langle x_i, x_j \rangle$, intersection kernel $K_{HI}(x_i, x_j) = \min\{x_i, x_j\}$ and the $\chi^2$ kernel $K_{\chi^2}(x_i, x_j) = 2\langle x_i, x_j \rangle/(x_i + x_j)$ [1].

3 SVM Multiclass Decomposition Rules

To solve a multiclass problem different solutions going from the decomposition of the multiclass problem until new formulations of the SVM optimization problem have been presented. One-versus-one (ovo) and one-versus-all (ova) are nevertheless the most common strategies. The ovo strategy constructs C(C − 1)/2 binary problems solving every possible class combination whereas the ova defines C hyperplanes to classify one class against all the other. In spite of the simplicity, both approaches define a region of ambiguity, meaning that when an instance falls in this region, its class cannot be predicted precisely. Take as example Fig. 1. Ascertaining patterns which can differentiate Fig. 1(b) from Fig. 1(c) may not be an easy task, especially for an inexperienced physician.3 In order to leverage these differences we tailored a learning strategy borrowing concepts from multiclass decomposition rules and prior knowledge from our problem [3]. Let us review these concepts in the first place.

NBI was presented as a reliable tool for a quick visualization of the mucosa. This technology allows the enhancement of pit patterns and microvasculature which evidence biological morphological details with histo- logical correspondence. Each pit pattern and microvascular structures originated the proposal of a taxonomy with 4 types. Each one represents a morphological pattern that is present in the Barrett’s segment when NBI endoscopes are used: Type A: Round pits with regular microvasculature;
Figure 1: Singh taxonomy [3] illustrative figures from our dataset. (a) Type A: Round pits with regular microvasculature; (b) Type B: Villous/ridge pits with regular microvasculature; (c) Type C: Absent pits with regular microvasculature; and, (d) Type D: Distorted pits with irregular microvasculature.

Type B: Villous/ridge pits with regular microvasculature; Type C: Absent pits with regular microvasculature; and, Type D: Distorted pits with irregular microvasculature (see Fig. 1) [3]. Each one of these types represent a degradation on the mucosa morphology: type A to B, we have without and low dysplasia, and from C to D we have low to high dysplasia.

Our proposal encompasses on a two-step learning process. A first, straightforward approach, consists on learning normal/metaplasia (say, \(\mathcal{AB}\)) from dysplasia cases (\(\mathcal{CD}\)). The second, and to solve the multiclass problem based in our previous analysis, since the discrimination of types B and C (see Fig. 1(b) and Fig. 1(c)) may be prone to error due to similar visual cues, we merged these two classes into a single one. This step thus consist on inferring decisions over simple cases (without meta-

B and C (see Fig. 1(b) and Fig. 1(c)) may be prone to error due to similar visual cues, we merged these two classes into a single one. This step thus consists on inferring decisions over simple cases (without meta-

<table>
<thead>
<tr>
<th>rwf</th>
<th>Vocabulary Size (# visual words)</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>twx</td>
<td>SDR</td>
<td>0.74 ± 0.03</td>
<td>0.80 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, sin. voc.</td>
<td>0.65 ± 0.03</td>
<td>0.67 ± 0.01</td>
</tr>
<tr>
<td></td>
<td>ovo, mul. voc.</td>
<td>0.68 ± 0.03</td>
<td>0.63 ± 0.02</td>
</tr>
<tr>
<td>bwx</td>
<td>SDR</td>
<td>0.41 ± 0.03</td>
<td>0.58 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, sin. voc.</td>
<td>0.32 ± 0.03</td>
<td>0.35 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, mul. voc.</td>
<td>0.29 ± 0.01</td>
<td>0.32 ± 0.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rwf</th>
<th>Vocabulary Size (# visual words)</th>
<th>100</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>twx</td>
<td>SDR</td>
<td>0.74 ± 0.03</td>
<td>0.80 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, sin. voc.</td>
<td>0.65 ± 0.03</td>
<td>0.67 ± 0.01</td>
</tr>
<tr>
<td></td>
<td>ovo, mul. voc.</td>
<td>0.68 ± 0.03</td>
<td>0.63 ± 0.02</td>
</tr>
<tr>
<td>bwx</td>
<td>SDR</td>
<td>0.41 ± 0.03</td>
<td>0.58 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, sin. voc.</td>
<td>0.32 ± 0.03</td>
<td>0.35 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>ovo, mul. voc.</td>
<td>0.29 ± 0.01</td>
<td>0.32 ± 0.03</td>
</tr>
</tbody>
</table>

Table 2: Accuracy results (± std. dev.) when using the linear kernel.

4 Experimental Study and Discussion

Our dataset consists of 250 images collected by a NBI endoscope, manually annotated by a clinical expert according Singh’s [3] taxonomy with clinical relevant regions identified. In our study we a two scale SIFT (\(4 \times 4\) and \(10 \times 10\)) was employed with sampling points separated by 10 pixels. \(txx\) and \(bxx\) term weighting functions as described in Section 2 were also used. Data were randomly split in two sub-sets: training and testing with 40% and 60% of the data respectively. The best parameterization for each model was found by a ‘grid-search’ based on a 3-fold cross validation construction were also analyzed: building upon the whole data (identified as sin. voc.) in contrast to a stratified approach which builds a vocabulary for each class (identified as mul. voc.). For a faithful comparison, we fixed the vocabulary size for the mul. voc. approach by setting equal portions so that the total number of visual words per class equals the desired vocabulary size. A first analysis from results depicted in Table 1 to Table 2 shows that our SDR outperformed the other decomposition rules attaining a performance rate of 80%. This result can be advantageous towards SDR once is requires to train only two classifiers in comparison to the \(C(C - 1)/2\) and \(C\) classifiers required for the ovo and ova, respectively. A more detailed analysis shows that the number of visual words within a vocabulary does not always guarantee better recognition rates (see Table 2 for the \(txx\)). This behavior is due to the scarcity of visual words on the histogram lowering the discriminative power of the vocabulary (contrast with \(bxx\) in Table 2).

5 Conclusion

Existing CAD systems for gastroenterology examinations using local features such as SIFT do not detail thoroughly the influence of each technique such as decomposition rules and term weighting approaches. In this work we studied the aspects that influence the recognition rates when using SIFT applied to gastroenterology images. In the end an overall result of 80% was attained by using our SDR coupled with a term frequency weighting function.

Acknowledgments

This work was partially financially supported by FCT (Portuguese Science Foundation) grant PTDC/EIA-CCO/109982/2009 and Pest-OE/EEI/LA0008/2013.

References


[3] R. Singh, GK Anagnostopoulos, K. Yao, H. Karageorgiou, PJ For-

Heart Sound Analysis for Cardiac Pathology Identification: Detection of Systolic Murmurs

João Pedrosa1,2
bio08014@fe.up.pt
Ana Castro1,3
ana.castro@dcc.fc.up.pt
Tiago T. V. Vinhoza1,3
tiago.vinhoza@ieee.org

Abstract
The digital analysis of heart sounds has revealed itself as an evolving field of study. Many attempts are made to create decision support systems capable of diminishing hospital costs and helping physicians in the first screening through algorithms capable of segmenting a phonocardiogram into its cardiac cycles and detecting and characterizing murmurs. The work proposed consists on the development of novel algorithms for the segmentation of the heart sounds into heart cycles as well as feature extraction and murmur detection and classification in a realistic clinical environment. The segmentation algorithm was tested returning a sensitivity and positive predictive value of 89,2% and 98,6%, respectively. The murmur detection was evaluated in two different situations, with a random division between train and test set and a division according to patients. The first returned sensitivity and specificity of 98,42% and 97,21%, respectively. The second division had a far worse performance with a minimum error of 33,65%. The operating point was chosen at sensitivity 69,67% and a specificity 46,91% for a total error of 38,90% by varying the percentage of segments classified as murmurs needed for a positive murmur classification.

1 Introduction
The heart, centre of the cardiovascular system, is responsible for pumping blood to every vital organ of the human body and its correct functioning is therefore of paramount importance. To fulfil its function, the heart follows a series of coordinated movements known as the heart cycle, which can be separated as the systole (muscle contraction) and the diastole (muscle relaxation). A distinctive sound, lub-dub, is produced by these movements, known as the phonocardiogram (PCG). This sound can however be modified by a number of cardiopathies, producing heart murmurs. These murmurs can be either pathological or innocent and arise differently depending on their origin. 50-70% of children have a murmur (most of these are innocent) and 2-5% of adults have or will eventually have a pathological valvular murmur [1-3].

Several tools exist in the clinical environment to confront the problem, such as auscultation with, however, some disadvantages. Cardiac auscultation through the use of the stethoscope is the most common technique however it is highly subjective and is dependent on a lengthy and continuous training. Echocardiography, despite its higher precision, is not available everywhere and is associated with higher costs and patient stress. Computer-aided auscultation would then allow faster and cheaper decisions by using a tool widely known by both physicians and patients [4, 5].

The main goal was then to design a robust algorithm capable of segmenting a PCG signal into heart cycles and heart sounds and also able to detect systolic. The designed algorithm should be independent of other sources of information such as electrocardiography or echocardiography and should be able to cope with the amount of noise and variability that exists in a real clinical environment.

2 State of Art
Computer-aided auscultation has been a subject of research for some time and many different methods have been applied to solve this problematic not only in terms of the proper algorithm and classification but also in the very way the data is acquired and treated. PCG signal processing can be crudely divided into two main research areas. One is focused in the detection of events such as S1 and S2 in order to perform a segmentation of the PCG into heart cycles. The other is the detection of murmurs and consequently of cardiac pathologies. However, because both objectives are interconnected and are accomplished from the same base signal, the PCG, both objectives share the same basic processing tools. These processing tools range, nevertheless, a large amount of techniques from the most simple to some utterly complex [4].

On the subject of heart sound segmentation, H. Liang et al. developed a method based solely on Shannon Energy envelopes obtaining sensitivity and positive predictive values of 94,11% and 98,76% respectively [6]. H. Naseri et M. R. Homaeinezhad used an approach based on the detection of high amplitude and frequency specific events using the Fast Fourier Transform. The sensitivity and positive predictive values obtained were of 99% and 98,6% respectively [7]. D. Gill et al. used self-organizing probabilistic maps to distinguish the peaks of a homomorphic envelope obtained from the PCG obtaining sensitivity and positive predictive values of 98,40% and 96,7% respectively [8].

Regarding murmur detection, C. Ahlstrom et al. used time domain and time-frequency features to train a classifier with 86% accuracy [9, 10]. D. Kumar et al. used the complexity signatures to obtain a murmur detection classifier of 91,09% sensitivity and 95,25% specificity. E. Delgado-Trejos conducted a wider analysis of the features using time domain, time-frequency domain, perceptual and nonlinear features obtaining an overall accuracy of 96,11% [11].

3 Methodology
The approach used is composed of two main phases. First the main heart sounds are identified so that the signal is separated into systolic and diastolic segments. Each of the segments is then subjected to a classifier to determine the presence of a murmur.

3.1 Heart Sound Segmentation
As the rest of the detection algorithm is dependent of a precise segmentation, this stage needs to be as robust as possible and not only identify the correct sounds but also correctly classify them as S1 and S2. The segmentation algorithm designed is mainly based in the autocorrelation function to find the periodic components of a signal.

First of all, a pre-processing routine is applied composed of continuous wavelet denoising using the Morlet mother wavelet with scales corresponding to frequencies of 15-150Hz. This mother wavelet was chosen as it has been proven to be the most effective for continuous wavelet transform of PCGs [12]. The remaining signal is converted to its energy envelope, normalized and decimated to obtain a smooth envelope. Secondly, the PCG is divided into 1,5s segments and a mean amplitude criterion is used to discard segments with high amplitude noise. The periodic elements of the remaining segments are then detected using the autocorrelation function and the systole length is estimated according to the function’s peaks. Finally, the heart sounds detected using the estimated systole length are classified into S1 or S2. While this is fairly easy in most cases as the systole is normally shorter than the diastole, in children and adults with higher heart rates systoles and diastoles are approximately of the same length. Non-duration based criteria must then be used such as the differences between the frequency spectra shape of the main heart sounds [13].

3.2 Feature Extraction and Classification
A total of 250 features were extracted from each segment to classify each segment as presenting murmur or not. These features were extracted from different analysis domains to ensure the segments were described as thoroughly as possible. Time domain analysis was chosen as it allows the study of the amplitude in the systole, which is altered in the presence of a murmur. A time-frequency analysis was also used to identify the correct frequency spectrum of the murmurs and distinguish them from noisy events. Perceptual analysis is a tool that perceives frequency in a logarithmic fashion as it is perceived by the human ear. It has been proved that murmurs are the highest complexity events in a PCG. This makes a nonlinear study of the PCG a promising tool for murmur detection, as the presence of a murmur will raise the overall complexity of a PCG [9, 11]. Table 3.1 shows the features extracted as well as its amount and analysis domain.
A k-means classifier was trained using these 250 features to classify the segments. A sequential forward feature selection (SFFS) algorithm was also used to determine the optimal feature set. Because the classifier would evaluate each systolic segment, a threshold was applied to the amount of segments classified as presenting murmur to define whether or not a patient was considered to present a murmur or not.

4 Results

To evaluate the performance of the designed algorithms two different databases were used. The DigiScope database is a relatively small database composed of 72 signals (~1min each) acquired in an unrestrained clinical environment with a physician-friendly protocol, as close as possible to the routine used in normal auscultation. The signals are then composed of the four main auscultation spots and have high levels of noise. The presence of murmurs and its temporal localization (systole/diastole) is annotated but the main heart sounds are not however. The patients composing this database are of ages from six months to 17 years. A secondary database was then used, a section of the PASCAL CHSC 2011 with labelled heart sounds. It is composed of 111 signals (1-30s) without murmurs and with very little or no noise. The patients include both children and adults [14].

The PASCAL database was then used to evaluate the performance of the segmentation algorithm. A sensitivity of 89.2% and a positive predictive value of 98.6% were obtained and an additional measure, the average temporal deviation of the detections, was also computed returning a value of 9.8 ms. These performance values are well within what is found in literature except for the imbalance found between the sensitivity and the positive predictive value. This is however not a problem as the goal of this algorithm is to detect murmurs and not to achieve a perfect segmentation. It is not important to find every heart sound but is extremely important that those found are in fact heart sounds. The 9.8 ms average deviation is a small deviation given that the average duration of the main heart sounds is of approximately 100ms.

However, this is the error that should be expected when classifying a new patient due to the patient variability that exists even in patients with the same pathology. This effect is augmented even further due to the fact that the database used is relatively small and due to the noise present in the signals used. The sensitivity and specificity of the trained classifier was also studied returning values at minimum error of 52.38% and 79.40% respectively. By changing the patient threshold the operating value was changed to a sensitivity of 69.67% and 46.91% specificity to ensure a greater amount of murmurs detected.

5 Conclusions and Future Work

This paper presented a novel algorithm for the digital analysis of heart sounds and detection of murmurs. Promising results were obtained for both the segmentation algorithm and the murmur detection. Given the motivation of this paper it is important to compare the obtained values to the accuracy levels of general practitioners. M. Lam et al. conducted a study that, among other things, evaluated the accuracy of physician trainees in the detection of murmurs obtaining accuracy values of 79.2% and 67% for pan-systolic and ejection systolic murmurs respectively [15]. By comparison to the obtained sensitivity value of 69.67%, one sees that it is within the range of the values obtained by M. Lam et al. for the physician’s normal performance. To make a difference in the clinical environment, a murmur detection algorithm should however have a superior accuracy to that of physical trainees. To achieve this, different features should be experimented to obtain an optimal feature set. Furthermore, we must enlarge the database to diminish the effect of interpatient variability.

Acknowledgments

The writers would like to acknowledge Dr. Sandra Mattos (Real Hospital Portugués, Brazil) and Prof. Miguel Coimbra (Universidade do Porto) for sharing the DigiScope database. This work was supported by FCT project PEst-OE/EEI/LA0008/2013.

References


Table 3.1- Features extracted for murmur detection.

<table>
<thead>
<tr>
<th>Analysis Domain</th>
<th>Feature Name</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Domain</td>
<td>Shannon Energy</td>
<td>12</td>
</tr>
<tr>
<td>Time-Frequency Domain</td>
<td>Continuous Wavelet Transform</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Discrete Wavelet Transform</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Singular Value Decomposition</td>
<td>48</td>
</tr>
<tr>
<td>Perception</td>
<td>Mel-Frequency Cepstrum Coefficients</td>
<td>100</td>
</tr>
<tr>
<td>Nonlinear and Chaos Based</td>
<td>Bispectrum</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Variance Fractal Dimension</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Lyapunov Exponents</td>
<td>2</td>
</tr>
</tbody>
</table>

Total: 250

Figure 4.1 – Murmur classification error for the six different feature sets tested. Random test and train set division according to patients.

The performance of the trained classifier was evaluated in two different approaches. The first approach was conducted by performing a random train-test division disregarding the patient to whom each segment originated. A minimum error of 4.65% was obtained using the whole feature set and a minimum error of 2.19% was obtained for the optimal subset of 167 features. These values are well within the range of those found in literature where the same test-train division was used. This approach does not however consider the interpatient variability as segments from the same patient are divided between the train and test set. This second approach divides the patients between the train and the obtained results are shown on Figure 4.1. The individual performance of each of the classes of features was also studied.

As expected, the error is much larger than in the previous division.
Forecasting the Usage of Home Appliances with Denoised Signal Patterns

Marisa Figueiredo¹
mbfig@dei.uc.pt
Bernardete Ribeiro¹
bribeiro@dei.uc.pt
Ana de Almeida²
Ana.Almeida@iscte.pt

¹ Department of Informatics Engineering
CISUC, University of Coimbra
Coimbra, Portugal
² Sciences and Technologies of Information Department
ISCTE-IUL University Institute of Lisbon
Lisboa, Portugal

Abstract
As the Internet of Things becomes reality and software appears in every device and in every building, we are interested to reach a level of ambient intelligence that promotes the understanding of our home real usage. Thus the need for forecasting electrical home consumptions to help changing human behavior towards energy saving for a sustainable environment. With this aim, in this paper we propose a Non-Linear AutoRegressive with eXogenous inputs (NARX) forecasting consumption model of an electrical aggregated signal. Support Vector Regression (SVR), which exhibits good generalization due to its regularization capability, was successfully tested on onsite real data measurements. The method yields good to excellent results when analyzing the Mean Squared Error and Squared Correlation Coefficient on the household collected test data.

1 Introduction
Electrical energy efficiency, the optimization of its consumption, can be driven by providing the appropriate feedback to the consumers. Among the useful information that several studies emphasize as important to accomplish consumption reductions, one can find detailed device consumption data. Regarding the enrichment of this type of feedback, the present work investigates time series forecasting of the electrical consumption of appliances that require human intervention to operate (henceforward denominated as non-automatic). Non-Linear AutoRegressive with eXogenous inputs (NARX) model is implemented using Support Vector Regression (SVR) [2] for predicting household electrical consumption. This model was evaluated using electrical signal patterns resulting from the Embedding Wavelet Shrinkage and Denoising (EWD) approach proposed by the authors in [3]. Due to the good generalization capability of the Support Vector Regression, the model yields excellent results regarding the mean squared error and squared correlation coefficient.

2 Time Series Forecasting of Home Usage Appliances
In the following, Support Vector Regression (SVR) and Non-Linear AutoRegressor with eXogenous inputs (NARX) models are briefly introduced. The SVR is no more than an extension of the Support Vector Machines (SVM) learning algorithm for regression. Both SVR and SVM are based on the statistical learning theory developed by Vapnik and Chervonenkis [5]. The usage of a nonlinear kernel transformation to map the input samples into a higher-dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive input samples into a higher-dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive input samples into a higher-dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive input samples into a higher-dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive input samples into a higher-dimensional space (feature space), where linear regression or classification can be performed, is the major distinctive

\[ f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x) + b. \]

For the nonlinear case, several kernel functions have been proposed which satisfy the Mercer’s condition (more details in [5]). The choice of a kernel, which depends on the application, is usually done by trial-and-error. Prior knowledge about the linearity of the data leads to a linear kernel or, in the non-linear case, it helps the kernel selection. The usual strategy for kernel selection consists in the observation of the prediction performance on validation data. The most common kernels are: linear, \((x_i, x - j)^2\); polynomial, \((x_i, x - j)^p\); sigmoid, \(\tanh(b(x_i, x - j))\) and Radial Basis Function (RBF) or Gaussian, \(\exp\left(-\gamma|\|x_i - x_j|\|^2\right)\).

Stochastic models able to predict future values using previous observations are a powerful tool for time series prediction. A general class of such models is known as the Non-Linear AutoRegressive Moving Average with eXogenous inputs (NARMAX). Such models are described as

\[ y(t) = F[y(t-1), \ldots, y(t-n_y), e(t-1), \ldots, e(t-n_e), x(t-1), \ldots, x(t-n_x)] + e(t), \]

where \(F\) is an unknown function, \(y, e\) and \(x\) are the output, noise and external input of the system model, respectively, and \(n_y, n_e\) and \(n_x\) are the maximum lags in the output, noise and input. Among the several special cases of the general NARMAX model, one of the most commonly used is the Non-Linear AutoRegressor with eXogenous inputs (NARX) model. This method models \(F\) in terms of \(y, e\) and \(x\), and \(n_y, n_e\) and \(n_x\) which are the corresponding maximum lags. In this work, a NARX model based on SVR (NARX-SVR) gives the approximation of \(F\) using SVR [4] for prediction of the future electrical energy consumption associated with appliances used at home, the non-automatic devices in particular.

3 Computational Experiment
The experiment designed to evaluate the proposed forecasting model whose input is the EWD denoised signal is described in the following.

3.1 Data Set Description
For the evaluation of the EWD and NARX-SVR performance we used home electrical consumption measured at a real household. The acquired signal represents the total consumption of the devices at the home: (i) the non-automatic and (ii) the ones, as the refrigerator, that once connected to the electrical circuit operate automatically. The data set consists of signals where each sample reports the average active power in the house electrical circuit over each 15 minutes during a day. The signals were collected from a given household during a four month period [1] and were divided in two sets: one corresponding to a six vacation day period, where no human activity interference was observed, and another containing load signals from 107 “regular” days, i.e., with the normal everyday usage of the household by its inhabitants.

3.2 The Aggregated Denoised Signals
The information associated with the automatic appliances was regarded as being the ‘noise’ in the measured signal. Thereby, the data set signals were denoised based on the author’s approach EWD (Embedding Wavelet Shrinkage and Denoising) illustrated in Figure 1. Different types of wavelet functions were tested for each one of the initial signals to choose the more adequate for each one of these signals. In order to save computation time and considering the results thus obtained, a subset of (the more adequate) wavelet functions was selected to be used.
EWD approach in accordance with previous experiments regarding the Wavelet Shrinkage performance for the task at hand (see [3]). The outcome signals are the input time series for this forecasting experiment.

### 3.3 Experimental Setup and Performance Metrics

The output signal from the EWD, i.e., the signal comprising the consumption of non-automatic devices was used as the input for this forecasting experience. The entire data set of denoised signals associated with the usual activity days in the house was divided in two sets: two thirds are used for the model training and the remaining signals are used for evaluating the generalization capability of the proposed model. For the NARX implementation, the lag values \( n_1 \) and \( n_2 \) were set to four and the \( x \) input represented the time instant of the day (in minutes) associated with the sampling index \( d_i = 1, \ldots, 96 \), for each day, i.e., \( x = 15 \times d_i \). Concerning the SVR, the RBF was the chosen kernel due to its well-known good performance. The free SVR parameters (\( \gamma, C \)) were then found throughout a grid search and the values used in the experiment were: \( \gamma \in \{0.001, 0.005, 0.01, 0.1, 1, 1.5, 10, 100\} \), \( C \in \{1, 5, 10, 25, 50, 75, 100, 150, 200\} \), and the parameter \( \varepsilon \) was set to 0.0001. To evaluate the forecasting model performance, Mean Squared Error (MSE) and Squared Correlation Coefficient (SCC) were computed. The former quantifies the difference between the target values \( y_i \) and the prediction \( f(x_i) \) by measuring the average of the squares of these differences. Formally, it is defined by

\[
\frac{1}{n} \sum_{i=1}^{n} ( f(x_i) - y_i )^2
\]

where \( n \) is the number of predicted observations. As for SCC, also usually denoted by \( r^2 \), it quantifies the correlation between two variables, namely, the target signal and the predicted one, according to the ratio

\[
\frac{(\sum_{i=1}^{n} f(x_i) - \bar{f}(x)) \cdot (\sum_{i=1}^{n} y_i - \bar{y})}{\sqrt{(\sum_{i=1}^{n} (f(x_i) - \bar{f}(x))^2) \cdot (\sum_{i=1}^{n} (y_i - \bar{y})^2)}}
\]

## 4 Results and Discussion

The MRE and SCC results\(^1\) search the values of the evaluation metrics using several test values for \( \gamma \) and \( C \) in the NARX-SVR model. It was observed, as expected, that the results are more favorable for the training phase than for the testing: the MRE (SCC) between \( f(x) \) and \( y \) is smaller (respectively higher) in the learning step than for testing. Regarding the training step and a fixed \( \gamma \) value, no relevant variation on the MSE (SCC) was observed when a increment of \( C (C > 1) \) occurred. Although, at test step, when \( \gamma = 0.001 \), a decrease (an increase, respectively) on the MRE (SCC) values was observed as \( C \) ranges from 5 until 25. For each fixed \( C \) value, as the value of \( \gamma \) is incremented in training, the SCC increased 29% (from 0.775 to 0.996) while the MRE decreases (e.g., from 0.2 to 0.008). A contrasting trend was observed for the test results. When \( \gamma \) decreases, the SCC values increase and accordingly, the MRE values decrease (from 0.807 to 0.205). In fact, an increase of 3.5 times was observed (from 0.229 to 0.797) when \( \gamma \) range from 100 to 0.001 for the test SCC values. In order to find a satisfactory tradeoff between the results of training and test and, at the same time, prevent overfitting, the search grid was refined around \( \gamma = 0.001 \) and \( C \in \{5, 10, 25\} \) using the most attractive results so far obtained. The SCC results are presented in Figure 2. For the test set, the smallest MRE values were yielded for the following settings: \( \{0.009, 25\} \) and \( \{0.001, 10\} \). Slightly better results (higher SCC and lower MRE) were achieved for the former. The test MSE is 2% lower for the parameters \( \{0.009, 25\} \) than for \( \{0.001, 10\} \) corresponding to a SCC of 0.801. As can be depicted from Fig. 3, despite of the efficient prediction for the training set, no overfitting occurs as shown by the good prediction of \( f(x) \) obtained with the test set.

## 5 Conclusions

The combined approach of denoised electrical signals and the NARX-SVR model for forecasting the electrical consumption of non-automatic appliances presents a very satisfactory performance. We intend to extend this study by considering a signal lag-1 forecasting classifier particularly in building early-warning systems for usage of home electrical appliances.

### Acknowledgments

FCT (Foundation for Science and Technology) is gratefully acknowledged for the grant SFRH/BD/68353/2010.

### References

Learning from uneven video streams in a multi-camera scenario

1 INESC TEC
Porto, Portugal
2 Faculdade de Engenharia
Universidade do Porto

Abstract

We present a semi-supervised incremental learning algorithm for evolving visual data in order to develop a robust and flexible track classification system in a multi camera surveillance scenario. Most existing methods, which are variations on static learning schemes, cannot cope with many real-life challenges. The scarcity of labelled data in real applications ends up generating poor classifiers. Furthermore, labelling the whole data (possibly massive in such applications) imposes a high cost to the system, rendering the technology impractical. So, there is an increasing interest on exploiting un-labelled data.

Our proposed method learns from consecutive batches by updating an ensemble in each time. It tries to achieve a balance between performance of the system and amount of data which needs to be labelled. As no restriction is considered, the system can address many practical problems in an evolving multi-camera scenario, such as concept drift, class evolution and various length of video streams which have not been addressed before.

1 Introduction

Over the last decades, video surveillance began to spread rapidly, specifically targeted at public areas. Recording for hours, days, and possibly years provides a massive amount of information coming from an evolving environment in which traditional learning methods fail to reflect the evolution taking place. In such environments, the underlying distribution of data changes over time - often referred to as concept drift - either due to intrinsic changes (pose change, movement, etc.), or extrinsic changes (lighting condition, dynamic background, complex object background, changes in camera angle, etc.). Thus, models need to be continually updated to represent the latest concepts. The problem is further aggravated when new objects enter the scene - referred to as class evolution in machine learning literature - as new models need to be trained for the novel classes.

Figure 1 demonstrates a typical surveillance scenario. Depending on the angle of the view and quality of the camera, every surveillance camera covers an area called Field of View (FoV). Often the fields of view are disjoint due to budget constraints, whereas they overlap in some scenarios. When entering the scene, the object will enter the coverage area of at least one of the cameras. The surveillance system will have to track that object from the first moment it was captured by a camera and across all cameras whose field of view overlaps the object’s path. In such environments where objects move around and cross the FOV of multiple cameras, it is more than likely to have multiple streams, potentially overlapping in time, recorded at different starting points with various lengths, for the same individual object (Figure 1).

Yet, much of the learning literature is concerned with a stationary environment, where fixed and known number of categories to be recognized [6] and enough resources (labelled data, memory and computational power) are available [4]. Learning from time-changing data has mostly appeared in data mining context and various approaches have been proposed. Ensemble-based approaches constitute a widely popular group of these algorithms to handle concept drift [1] and in some recent works class evolution [3], as well. Learn++.NSE [3] is one of the latest ensemble-based classification methods in literature, that generates a classifier using each batch of training data and applies a dynamic weighting strategy to define the share of each ensemble in the overall decision. As success is heavily dependent on labelled data, this method would not be applicable in wild scenarios. Although a considerable body of research has emerged that views whole or segments of a stream as a unique element to cluster (or classify), is a less explored area. The methods that have been proposed [2], require equal length streams coming from a fixed number of sources. Thus, they would fail to leverage information from time-varying video tracks. An effective and appropriate algorithm to fit in our scenario is required to: a) learn from multiple streams; b) mine streams with various length and starting points (uneven streams); c) handle the concept drift; d) accommodate new classes; e) deal with partially labelled or unlabelled data; f) be of limited complexity; g) handle multi-dimensional data. To the extent of our knowledge, no such method has been introduced in the literature.

2 Learning algorithm

In this section we present our Never Ending Visual Information Learning (NEVIL) framework. NEVIL is designed for non-stationary data environments in which no labelled data is available but the learning algorithm is able to interactively query the user to obtain the desired outputs at carefully chosen data points. The NEVIL algorithm is an ensemble of classifiers that are incrementally trained (with no access to previous data) on incoming batches of data, and combined with some form of weighted majority voting. A sketch of the proposed method is shown in Figure 2. A typical tracking algorithm analyses sequential video frames and outputs the movement of targets between the frames, generating multiple streams of visual data. Environmental challenges such as varying illumination, learning from multiple streams in wild environments,

Figure 2: NEVIL High-level Overview
nation, lack of contrast, bad positioning of acquisition devices, blurring caused by motion as well as occlusion make data often noisy and/or partially missing. We address these challenges by a batch divisive strategy, as learning from a data batch may reduce the noise and fill the gap caused by miss-tracking.

The algorithm is provided with a series of data batches $\mathcal{D}_m^t$, where $m$ is the index of the $i$-th stream present at time slot $t$, $TS_i$, (not all streams are necessarily present). Note that a stream corresponds to a track generated by the tracking system and a single camera can yield multiple streams. A single batch aggregates $B$ frames. The starting time of each stream is potentially different from stream to stream but batches are aligned between streams. Inside each frame the data corresponds to some pre-selected object representation (e.g. bag of words, histogram) and is out of the scope of this paper.

The ensemble obtained by all models generated up to the current time slot $TS_i$ is named the composite hypothesis $H_{t-1}$. With the arrival of the current data batches $\mathcal{D}_m^t$, $i = 1 \cdots M$, NEVIL tries to predict the class label for each of the batches in time slot $TS_i$ based on the probability estimate $p(C_k|\mathcal{D}_m^t, H_{t-1})$, where $C_k$ runs over all the class labels observed so far.

This kind of on-line learning approach addressed in this work can suffer if labelling errors accumulate, which is inevitable. Unrelated objects will sooner or later be assigned the same label or different labels will be assigned to different views of the same object. To help mitigate this issue, we allow the system to interact with a human, to help it stay on track. Initially, the composite model is initialized to yield the same probability to every possible class (uniform prior). When the batches $\mathcal{D}_m^t$ in time slot $t$ become available, NEVIL starts with computing the probabilities $p(C_k|\mathcal{D}_m^t, H_{t-1})$ for each batch $\mathcal{D}_m^t$ in the time slot. Once $p(C_k|\mathcal{D}_m^t, H_{t-1})$ are obtained, a batch confidence label (BCL) is estimated; if BCL is high (above a prespecified threshold), the predicted label

$$\arg\max_{C_k} p(C_k|\mathcal{D}_m^t, H_{t-1})$$

is accepted as correct, otherwise the user is requested to label the data batch. The labelled batches (either automatically or manually) are used to generate a new multiclass classifier that is integrated in the composite model, yielding $H_{t-1}$.

3 Experimental Methodology

3.1 Datasets

A series of experiments were conducted to explore the capabilities of the proposed framework. We tested the NEVIL framework with real video data from CAVIAR dataset [5] including: OneLeave ShopReenter1, Entry Exit Crossing Paths 1, OneShopOneWait1, OneStop Enter2 and WalkBy Shop1Front. Due to the presence of different perspectives of the same person, streams are drifting in time (see Fig. 3). These sequences present challenging situations with cluttered scenes, high rates of occlusion, different illumination conditions as well as different scales of the person being captured. We employ an automatic tracking approach to track objects in the scene and generate streams of bounding boxes, which define the tracked objects’ positions. As the method may fail to perfectly track the targets, a stream often includes frames of distinct objects [7]. An hierarchical bag-of-visterms method is applied to represent the tracked objects, resulting in a descriptor vector of size 11110 for each frame.

![Figure 3: An example of diversity in appearance](image)

3.2 Evaluation Criteria

Active learning aims to achieve high accuracy using as little annotation effort as possible. Thus, a trade-off between accuracy and proportion of labelled data can be considered as one of the most informative measures. Let $N$ denotes the total number of batches, $MC$ refers to misclassified batches, then the accuracy of the system in a given time slot is formulated as ratio of correctly classified batches to total number of batches. The total accuracy of a system over a period of time is derived from the mean accuracy of all the time slots. The Annotation Effort is formulated as ratio of manually labelled batches to the total number of batches. One expects that the accuracy increases with the increase of annotation effort.

3.3 Results

In our framework, Support Vector Machine (SVM) is applied as base learner. The size of the batches are optimized for each dataset. Figure 4 presents the results obtained from various CAVIAR sequences. In the OneStopMoveEnter1, the most complex scenario with 42 streams from 14 classes, NEVIL achieves 80% accuracy with 30% annotation effort.

4 Conclusion and Future work

In this paper, we proposed a semi-supervised incremental learning framework (NEVIL) to mine visual data coming from non-stationary environments. Inspired from active learning strategies, an oracle provides labelled batches. We have empirically shown that NEVIL can achieve high accuracy with fairly little human effort.

Although, framework fulfills most of the characteristics listed in Sec. 1, as the size of ensemble is growing in time, complexity is still the main concern and main direction for future research.

Acknowledgements

This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia within project FCOMP-01-0124-FEDER-037281.

References


Temporal subsampling impact on echocardiography based analysis of the left ventricle dynamics

Susana Brás
susanabras@ua.pt
José Ribeiro
jri@chvng.min-saude.pt
Augusto Silva
augusto.silva@ua.pt
José Luís Oliveira
jlo@ua.pt

1 Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA), DETI
Universidade de Aveiro
Aveiro, Portugal
2 Serviço de Cardiologia, Centro Hospitalar Gaia Espinho
Vila Nova de Gaia, Portugal

Abstract

The automatic classification of medical images depends of many factors such as the modality, image features, classification techniques, the existence of annotated corpora and the images’ quality. In this work we evaluated the impact of echocardiograms temporal resolution on the automatic detection of the left ventricle displacement. To access this information it was built a variation model for the distinct cardiac phases, and evaluated in original and several down-sampled representations. The results pointed out that lowest frame rates (14 and 10) have a negative impact on the displacement detection. Nevertheless, when the frame rate is reduced to 21 frames/sec, the impact does not compromise the results.

1 Introduction

The automatic characterization of the cardiac individual conditions becomes a main concern since the cardiovascular diseases lead the major cause of death worldwide [3]. Therefore the quantification of the left ventricular activity is of utmost importance, due to characterization of myocardial function [1]. In particular, the left ventricle activity is responsible for the body oxygenation [1]. One option in the analysis of the left ventricle is the echocardiography images. Methods that evaluate the left ventricle variation through a cardiac cycle help in the automatic discrimination between groups (normal versus pathological). However, depending on the machine used there is different frame rate [2]. Sometimes, it is needed to reduce the echocardiogram frame rate, because higher frame rates occupy a huge space on storage. However, it is mandatory to guarantee the data quality for posterior analysis.

The correct analysis of a signal is dependent on its sample frequency. A perfect reconstruction is only possible if the Nyquist theorem is accomplished, which states that the sampling frequency should be greater than twice the maximum frequency of the signal. Artificial mechanisms to increase the sample frequency are not enough to balance the lack of information. In this paper, it is presented a study evaluating the variation of the left ventricle due to the physiology.

2 Dataset

The echocardiography images were collected in regular procedures at Centro Hospitalar Gaia Espinho. The local ethics committee approved the procedures and an informed consent was provided to the patient. In all collected data, a four chamber view was recorded during, at least, a complete cardiac cycle. The data were then anonymized and offline processed using a proprietary software from Siemens, the syngo Velocity Vector Imaging technology (VVI) [4]. The left ventricle contour is visually identified and manually marked, at the beginning of the systolic phase. This first landmark is not automatically adjusted to the image. However, the software is able to track the evolution of the first landmark visually identified and manually marked, at the beginning of the systolic phase. This first landmark is not automatically adjusted to the image. However, the software is able to track the evolution of the first landmark. In this work we evaluate the impact of echocardiograms temporal resolution on the automatic detection of the left ventricle displacement. To access this information it was built a variation model for the distinct cardiac phases, and evaluated in original and several down-sampled representations. The results pointed out that lowest frame rates (14 and 10) have a negative impact on the displacement detection. Nevertheless, when the frame rate is reduced to 21 frames/sec, the impact does not compromise the results.

3 Methods

The echocardiogram frame rate may vary between 10 to 80 frames/sec [2]. VVI follows the landmark through consecutive frames, estimating the left ventricle contour in a range of 49 points (delineating the ventricle).

In the present study, the echocardiogram with highest frame rate in our database was chosen (43 frames/sec). For each point of the 49 data points, it will estimate (approximately) 55 positions through one cycle (cycle duration of 1.3 seconds). For data analysis, the frame rate was decreased after acquisition, by downsampling to achieve a 21, 14 and 10 frames/sec. For each of the four models, it was built an approximation, based on a cubic spline, allowing to extrapolate the initial 55/27/18/13 points, associated with frame rates of 43/21/14/10, to 127 positions (100 frames/sec). This approach allows the comparison between the echocardiograms, guaranteeing the same number of frames to analyse in each second. The chosen echocardiogram represents two complete heart cycles.

3.1 Smooth Cubic Spline

The spline will find a cubic polynomial between two xx points and linking them, being its function defined as: \( f(x) = P_i(x), x_i \leq x \leq x_{i+1}, i = 1,...,k-1 \), where \( k \) is the length of the xx vector. The spline building is based on a balance between the best data fitting and the best smoothing. Therefore, to obtain the best spline to the data, eq. (1) is minimized [5].

\[
(1-q) \sum_{i=1}^{k} (y_i - f(t_i))^2 + q \int_0^1 f''(t)^2 dt
\]

where \( q \) is a parameter that varies between 0 and 1, and \( y \) is the "real" data.

3.2 Left Ventricle Contour variation Estimation

The cardiac cycle is divided in systole and diastole. In the systole the ventricle decreases its volume until achieve the minimum. In the diastole, the ventricle increases its volume until its maximum. Consequently by the volume curve inspection, the cardiac phases are identified. After the spline building and the data interpolation to 127 points, and considering this particular example, the data (xx and yy contour coordinates) were split in four phases (two systole and two diastole). In each of them, mean and standard deviation of interval coordinates were calculated, allowing the construction of a variation model from the ventricle movement in each cardiac phase, represented as \( \text{mean} \pm 2\text{std} \).

3.3 Quantification of variation

To evaluate the level of variation, it is calculated the Euclidean distance (eq. 2) between the variation model (\( \text{mean} \pm 2\text{std left ventricle contour} \)) in each phase, to each tested frame rate.

Given two vectors \( a \) and \( b \), the Euclidean distance is calculated through:

\[
\text{dist} = \sqrt{\frac{n}{\sum_{i=1}^{n} (a_i - b_i)^2}}
\]

where \( n \) is the vectors length.

4 Results and Discussion

Figure 1 represents the left ventricle variation model (43 frames/sec) in two cardiac phases during two complete cardiac cycles (4 variation models). By the figure inspection, it is observed that in the diastolic phase (second and fourth representations), the ventricle area is larger than in the systolic phases (first and third), in accordance to the physiology. The same analysis was done in other considered frame rates. After the variation model creation, the distances between cardiac phases in...
Figure 1: Left ventricle variation model (mean ± 2std) representation through two complete cardiac cycles, describing systolic (first and third representations) and diastolic (second and fourth representations). Below, it is represented the left ventricle volume, with red marks representing the transition between each cardiac phase. tx/ty is the xx/yy left ventricle coordinates, respectively.

Table 1: Left ventricle contour variation associated with the downgrade in each cardiac phase. The negative values correspond to a ventricle displacement variation to the exterior of the benchmark (43 frames/sec).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Systole</td>
<td>Diastole</td>
<td>Systole</td>
<td>Diastole</td>
<td>Systole</td>
<td>Diastole</td>
</tr>
<tr>
<td>Systole</td>
<td>0%</td>
<td>-0.13%</td>
<td>-3.03%</td>
<td>0.21%</td>
<td>0%</td>
<td>0.09%</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>0%</td>
<td>-0.79%</td>
<td>2.70%</td>
<td>—</td>
<td>0%</td>
</tr>
<tr>
<td>Systole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>-0.47%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>0%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Systole</td>
<td>0%</td>
<td>2.40%</td>
<td>33.72%</td>
<td>6.15%</td>
<td>0%</td>
<td>-7.39%</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>0%</td>
<td>5.46%</td>
<td>2.51%</td>
<td>—</td>
<td>0%</td>
</tr>
<tr>
<td>Systole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>8.49%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>0%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Systole</td>
<td>0%</td>
<td>0.25%</td>
<td>9.52%</td>
<td>3.20%</td>
<td>0%</td>
<td>2.54%</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>0%</td>
<td>0.04%</td>
<td>26.88%</td>
<td>—</td>
<td>0%</td>
</tr>
<tr>
<td>Systole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>2.36%</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Diastole</td>
<td>—</td>
<td>—</td>
<td>0%</td>
<td>0%</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

same model) were calculated following eq. 2. Those distances give information about the displacement of the left ventricle contour, and it is not known which part of the variation is due to physiology or to downgrade. It was assumed that the model associated with the maximum resolution (43 frames/sec) correspond to the "true" physiologic alterations, i.e. our benchmark. Therefore, the differences in the same cardiac phase between the benchmark and other frame rate models will correspond to the variation produced by the downgrade of the image resolution. Table 1 represents the percentage of variation associated with downgrade, calculated by the ratio between the previous defined difference and the benchmark variation in the analyzed cardiac phase.

By the table inspection, it is observed that the cardiac phase most affected by the downgrade is the systolic phase. When the image resolution is reduced to 14 frames/sec, it is observed the worst results, in both xx and yy coordinates, with a variation due to downgrade of approximately 40%. Nevertheless, the minimum resolution model follows the benchmark with a lower error than when 14 frames/sec are used; however, with variation values due to downgrade of approximately 20%/30%.

5 Conclusion
This study evidenced that the image resolution is of upmost importance when the left ventricle contour is studied in terms of displacement. However, our results evidenced that reducing the acquisition resolution to 21 frames/sec does not compromise the evaluation of the myocardium function.

6 Acknowledgements
This work was supported by the European Regional Development Fund (FEDER) through the programme COMPETE and by the Portuguese Government through the FCT. Fundação para a Ciência e a Tecnologia, in the scope of the projects FCOMP-01-0124-FEDER-022682 (ref. PEst-C/EEI/UI0127/2011), Incentivo/EEI/UI0127/2013 (Instituto de Engenharia Electrónica e Telemática de Aveiro, IEETA/UA, Aveiro, www.ieeta.pt), and Cloud Thinking (CENTRO-07-ST24-FEDER-002031, co-funded by QREN, "Mais Centro" program). S. Brás acknowledges the postdoctoral grant by Cloud Thinking Project (ref. BPD/DETI/5259/2013).

References
Knowledge on Heart Condition of Children based on Demographic and Physiological Features

Pedro Ferreira\textsuperscript{1,4}, Tiago T. V. Vinhoza\textsuperscript{2,4} and Ana Castro\textsuperscript{2,4} \{pedroferreira,ana.castro\}@dcc.fc.up.pt, tiago.vinhoza@ieee.org
Felipe Mourato\textsuperscript{3}, Thiago Tavares\textsuperscript{3} and Sandra Mattos\textsuperscript{3} \{felipe.a.mourato,thicow\}@gmail.com, ss.mattos@cardiolbr
Inês Dutra\textsuperscript{1,4} and Miguel Coimbra\textsuperscript{1,4} \{ines.mccoimbra\}@dcc.fc.up.pt
\textsuperscript{1}CRACS-INESC TEC LA, Porto, Portugal
\textsuperscript{2}Instituto de Telecomunicações, Porto, Portugal
\textsuperscript{3}Real Hospital Português, Recife, Pernambuco, Brazil
\textsuperscript{4}Departamento de Ciência de Computadores, Faculdade de Ciências da Universidade do Porto, Porto, Portugal

Abstract

We evaluated a population of 7199 children between 2 and 19 years old to study the relations between the observed demographic and physiological features in the occurrence of a pathological/non-pathological heart condition. The data was collected at the Real Hospital Português, Pernambuco, Brazil. We performed a feature importance study, with the aim of categorizing the most relevant variables, indicative of abnormalities. Results show that second heart sound, weight, heart rate, height and secondary reason for consultation are important features, but not nearly as decisive as the presence of heart murmurs. Quantitatively speaking, systolic murmurs and a hyperphonetic second heart sound increase the odds of having a pathology by a factor of 320 and 6, respectively.

1 Introduction

Children are usually thought of as having healthy hearts. Therefore it may be a surprise to many people to learn that, in the US, nine out of every 1000 babies are born with a congenital heart abnormality. It is estimated that one third of these babies require intervention to prevent death in the first year of life. Approximately 1.3 million people living in the US today were born with a congenital heart defect, and at least half of these individuals are under age 25\textsuperscript{1}. In Portugal, the number of cardiac surgeries in children performed a year is around 500 and eight out of 1000 babies are born with a heart abnormality\textsuperscript{2}. In Brazil, place where we collected our data it is estimated that between eight and ten children out of 1000 are born with a congenital cardiac disease [1].

We study a population of 7199 children between 2 and 19 years old, from the northeastern part of Brazil. Our goal is to identify variables that may be more indicative of normality or pathology and use this information to build classifiers that can, in an automatic fashion, distinguish between normal and pathological cases. Such classifiers, and an increased understanding of the relations between physiological and demographic variables, may help on the decision making process, avoiding missing pathological patients when they are consulting with a less experienced professional. This would allow detecting and initiating the treatments earlier, improving patient outcome and reducing costs. Few works in the literature report on prediction of heart diseases using machine learning techniques. The University of California at Irvine (UCI) repository has some datasets related to cardiology. The one most related to our dataset is the “Heart Disease” [2]. According to the UCI website, this database contains 76 attributes, but all published experiments refer to using a subset of 14 of them. Experiments with the Cleveland database have focused on attempting to distinguish presence of disease from absence of disease and focused mainly on classification performance. Some of them use feature selection to build classifiers but do not focus on the individual feature importance as we do in this work. The results obtained with an instance-based learning algorithm (IB1) [3], a neural network algorithm [4], and with Radial Basis Function Networks [5] report an accuracy of 75.7% (±0.8), 87.5%, and 84% respectively. Previous work on a smaller dataset of 169 children [6], and performing an exhaustive search for the best classifier, produced an accuracy of 90.5%. For this dataset with 7199 children, our training accuracy is 93% when predicting cardiac pathological cases, using a SVM classifier.

2 Materials and Methods

2.1 Data

The data used in this study was collected in the Real Hospital Português (RHP), Brazil.

In the original dataset containing 7603 instances and 33 features, we performed pre-processing tasks, namely, data cleaning, data transformation, data normalization, as well as removal of irrelevant features, like ID, name of doctor, etc. Since body mass index (BMI) is used in our study and this feature is usually assessed for 2-year old or older children, we removed from our dataset all the children younger than 2 years old. Therefore, our dataset, after pre-processing, consists of 7199 cases and 17 features\textsuperscript{1}. Of these 7199 cases, 2507 (34.8%) are pathological, while 4692 (65.2%) are healthy. The main goal of this work is to apply feature importance metrics to rank the variables that are more predictive of cardiac pathogens.

2.2 Methodology

We take two approaches to find the most relevant attributes to predict pathologies. First, we use model independent metrics such as mutual information and chi-squared tests. Then, we use model specific metrics, namely a variable importance measure by a random forest classifier and a logistic regression odds ratio analysis.

The mutual information approach to feature selection consists in computing the mutual information between each attribute and the class variable. This gives information on how correlated they are. Mutual information is defined as $I(X;Y) = H(X) - H(X|Y)$, where $X$ is our class variable and $Y$ is the feature under analysis [7]. The intuition this result gives us is the following: If $X$ and $Y$ are statistically independent random variables, then $H(X;Y) = H(X)$, that is, the knowledge of $Y$ does not reduce the uncertainty about $X$. If the knowledge of $Y$ removes all the uncertainty about $X$, then $H(X;Y) = 0$, as a result $I(X;Y)$ is maximum. So, the greater the value of $I(X;Y)$ is, the more important is the feature according to this measure.

The chi-squared test is used as a test of independence between two random variables. The first step is to calculate the chi-squared test statistic, $\chi^2$, which resembles a normalized sum of squared deviations between observed and expected frequencies. The second step is to determine the degrees of freedom, $d$, of that statistic, which is essentially the number of frequencies reduced by the number of parameters of the fitted distribution. In the third step, $\chi^2$ is compared to the critical value of no significance from the $\chi^2$ distribution, which in many cases gives a good approximation of the distribution of $\chi^2$.

We then analyze the model-specific metrics. First, we calculate the variable importance as measured by a random forest classifier, then we apply a logistic regression and perform odds ratio analysis to infer the importance of each feature in the odds of having pathology.

3 Results and Discussion

In this section we present the results obtained by applying the feature importance metrics presented in the previous section to our database. The mutual information results, using the 17 features, place murmur as the feature that most reduces the uncertainty of the class variable (either pathological or not). In fact, murmur presents a score much higher than the remaining features (Table 1).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murmur</td>
<td>0.61</td>
</tr>
<tr>
<td>Secondary Reason</td>
<td>0.10</td>
</tr>
<tr>
<td>Weight</td>
<td>0.09</td>
</tr>
<tr>
<td>Primary Reason</td>
<td>0.05</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 1 – Mutual Information applied to the 17 features: 7199 cases

1 Height, weight, sex, age range, BMI percentile, systolic/diastolic blood pressure, blood pressure assessment, murmur, second heart sound, pulses, heart rate, disease history 1/2, primary/secondary reason, pathology (classPosition).
From the presented result, murmur plays an important role when predicting cardiac pathology, which is in accordance to clinical assessment [8]. Nonetheless, from all the cases in the database we observe that 5,000 patients have absent murmur, and of those, 404 (â¥%) have cardiac pathology. Since murmur seems to be determinant in cardiac pathology detection, to study the impact of the absence of this characteristic, and to evaluate which other variables may aid in the detection of these pathological cases, we apply mutual information to the dataset containing only the patients with absent murmur. Results are presented in Table 2.

### Table 2 – Mutual Information applied to the 17 features: 5000 cases with absent murmur

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.17</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>0.05</td>
</tr>
<tr>
<td>Height</td>
<td>0.04</td>
</tr>
<tr>
<td>Second Heart Sound</td>
<td>0.04</td>
</tr>
<tr>
<td>Secondary Reason</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Although weight, heart rate, height, second heart sound (S2) and secondary reason, reach the top of the ranking (Table 2), their relevance is lower when compared to murmur (Table 1).

Chi-squared results using the 17 features from the 7199 instances reinforce also the idea that murmur is a key factor when assessing a cardiac disease. When applying chi-squared test to the dataset with absent murmur, the feature ranking obtained is very similar to the result rankings obtained in the mutual information approach, and as referred before.

Moving to classifier-based approaches, we first compute the mean decrease Gini achieved by a random forest classifier to the 17 features from the 7199 instances. It focuses on measuring the total decrease in node impurities from splitting on the variable, averaged over all trees. A variable that decreases the Gini index the most is responsible for a decreased node impurity, hence it is the most important in terms of separating the target classes. Analyzing the results with 17 features from the 7199-case dataset, murmur is the crucial feature in order to correctly separate the target classes, as it decreases the Gini index in an approximate score of 1976 (Table 3).

### Table 3 – Variable importance as measured by a Random Forest classifier using all 17 features: 7199 cases

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean Decrease Gini</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murmur</td>
<td>1975.98</td>
</tr>
<tr>
<td>Secondary Reason</td>
<td>216.44</td>
</tr>
<tr>
<td>Weight</td>
<td>189.82</td>
</tr>
<tr>
<td>Height</td>
<td>172.65</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>149.61</td>
</tr>
</tbody>
</table>

We noticed that secondary reason is ranked in the top 2 twice and in fifth place two times. This can be explained by the fact that one of the possible secondary reasons is “presence of murmurs”. From the 881 occurrences of that label, 635 (â¥%) are associated with a cardiac disease. This reinforces the importance of murmur in the accurate detection/classification of cardiac pathology.

We then apply logistic regression and compute the odds ratio of the features. The most important feature according to the odds-ratio is the presence of systolic murmurs, which increases the probability of having pathology by a factor of approximately 320. The results of the logistic regression also show that having an hyperphonetic S2, increases the odds of having a pathology by six.

Finally, we performed a small experiment using a SVM classifier applied to the 7199 cases to predict cardiac pathologies using/not using murmur as feature. The results, obtained with 10 times 10-fold cross-validation, are in Table 5 and are consistent with the results obtained in the feature importance analysis. We evaluate the accuracy, sensitivity, and specificity.

### Table 5 – Predicting Pathology

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Using murmur</th>
<th>Not using murmur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>93.21</td>
<td>73.12</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.85</td>
<td>0.37</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.98</td>
<td>0.92</td>
</tr>
</tbody>
</table>

## 4 Conclusions

In this study we present an exploratory analysis of various cardiac and demographic features collected from children (with and without cardiac pathology), in standard clinical practice. The most important result that is drawn from the several techniques presented is the importance of the presence of murmur in the detection of cardiac pathology. Although this information is not new, and is widely used in clinical practice to assess cardiovascular state in conjunction with other demographic data [8], it should be noticed that when this feature is not present, the remaining variables analyzed in this study do not contribute as decisively to the pathology detection. Hence it is crucial to have accurate information on murmur presence.

### Acknowledgements

This work was partially funded by the DigiScope (PTDC/EIA-CCO/100844/2008) and ABLe (PTDC/EEI-SII/2094/2012 (FCOMP-01-0124-FEDER-029010)) projects and by the Fundação para a Ciência e a Tecnologia (FCT/Portugal) project Pest-OE/EEI/LA0008/2013.

### References

Abstract

In this paper we propose the use of textons to perform the segmentation between land and water in SAR images. Several texton-based methods have been applied in different fields of image processing, namely image classification, where they achieved good results. However, textons have never been tested as a way to perform segmentation in SAR imagery. In this paper the method is applied to several images of the Tagus river taken under different conditions, and is shown to achieve up to 94% of correct pixel classifications in the segmented image.

1 Introduction

This paper aims at developing an automatic method for the segmentation between land and water in SAR (Synthetic Aperture Radar) images, with the purpose of improving floodplain inundation models on the lower Tagus river, so as to help reduce their uncertainty and validate hydraulic flood models.

Even though SAR images have several advantages over regular satellite imagery (for instance, they can be acquired during the night and regardless of cloud cover), they also have several drawbacks, such as the difficult processing due to the high noise levels, lack of color information and the presence of artifacts caused by the acquisition process. This leads to an increased difficulty of using automatic methods to analyze these images and to the need of using robust techniques when working with them. Since this kind of images are very textured, a good way to perform segmentation on SAR images is by exploring the image’s texture data. This allows for robust segmentation even in noisy situations, since it does not use pixel intensities directly.


The method developed in this paper to perform segmentation explores image texture and is based on the texton method introduced by Malik et al. [5], which has been applied in different fields of image processing, namely image classification, but has not been used before with SAR images. The texton method will be explained in section 2.

2 Methodology

The definition of textons as they are used in this paper was introduced by Malik et al. [5]. Textons are defined as the cluster centers of the filter responses when a filter bank is applied to an image. A filter bank is applied to the image to be segmented, and each pixel can then be represented in a higher dimensional space by the filter responses on that pixel, that is, the pixels are now vectors where each position is a filter response of said pixel. These vectors are then clustered with the K-means algorithm, and the resulting centers are called textons. The algorithm we used is based on the work by Varma and Zisserman [6], and consists of three stages: the generation of a texton dictionary, the generation of texton histograms, and pixel classification, where they achieved good results. However, textons have never been tested as a way to perform segmentation in SAR imagery. In this paper the method is applied to several images of the Tagus river taken under different conditions, and is shown to achieve up to 94% of correct pixel classifications in the segmented image.

2.1 Stage I - Texton dictionary generation

Firstly, given an image to be segmented, some training areas must be selected. At least one area for each class (land and water) is needed, and the areas should be representative of each region’s textural characteristics.

Then, for each class, a number of randomly placed patches will be selected from the previously chosen training areas, the probability being dependent of the training area size. All further processing at this stage will be applied to each one of these patches.

After the selection of the training patches, the first step is to apply a filter bank to those patches. This way, each pixel of the patch will be represented by a point in a higher dimensional space, the dimensionality being equal to the number of filters in the filter bank. Each coordinate of these points is the response output of each of the filters that compose the filter bank, centered on that pixel. Varma and Zisserman [6] performed several tests on different filter banks (although they were being used for image classification, and not segmentation), and concluded that the best performing filter bank was the MR8 (Maximum Response 8), since it allowed for the best rate of correct image classification. It is thus appropriate to use this filter bank as a starting point for this work.

This filter bank can be seen in Figure 1 and consists of 38 filters: an edge and a bar filter, both at three scales and six orientations, a Gaussian and a Laplacian of Gaussian. However, only 8 filter responses are used, since for each scale of filters, only the maximum response orientation is used. This means that the filter bank is rotationally invariant, since both the Gaussian and the Laplacian of Gaussian have rotational symmetry.

![Figure 1: The MR8 filter bank. (a) Edge filters; (b) Bar filters; (c) Gaussian and Laplacian of Gaussian.](image)

After applying the filter bank to each patch, the vectors of filter responses for each class are clustered using the K-means algorithm. In this algorithm each vector is associated to a cluster of vectors, based on which cluster has its mean closest to the vector. The result of this step is a set of vectors that are representative of the filter responses of an image - these are called textons. The textons from all training areas are stored in a texton dictionary, that will later be used. This step is represented in Figure 2.

![Figure 2: Stage 1 - Generating the texton dictionary](image)
2.2 Stage II - Texton histogram generation

In this second stage, shown in Figure 3, the training images will be processed again, but with a different purpose. Since there is now a texton dictionary - a set of textons extracted from the training images - it is possible to check, for every pixel in each patch, what texton lies closest to it in the higher dimensional filter response space, using the Euclidean distance. For this purpose, the filter bank is applied to the patches and then, for every response vector, the closest texton is found. As a result, each pixel will be assigned to one of the textons in the dictionary.

Now we generate, once again for each training patch, a texton histogram, where the frequency with which each texton occurs in the patch is mapped. These histograms are used as a model for the training patches, as they define their texture characteristics. The histograms are normalized to sum to unity.

2.3 Stage III - Pixel classification

This is the final step, described in section 2.2, but instead of processing only a few patches, the number of patches to be processed corresponds to the number of pixels in the image. For each pixel to be classified, a patch is generated around it, to which the filter bank is then applied. After this process, the texton histogram for the patch is generated and compared to all the learned histograms of the previous step using the $\chi^2$ distance. Then, the pixel is classified as belonging to the class that the majority of the three closest histograms belong to, which corresponds to the application of the K-NN (K-Nearest Neighbours) algorithm using $k=3$.

3 Results

The developed method was implemented in MATLAB and was applied to three different images of the Tagus river, each measuring 1300 by 1100 pixels. The images, with pixel dimension 12.5 meters, were acquired by ERS (European Remote Sensing) and ASAR (Advanced Synthetic Aperture Radar) European Space Agency SAR sensors, under different conditions such as wind speed or rainfall intensity, which causes the appearance of water to change significantly. As such, they provide a good way to test the results such as windspeed or rainfall intensity, which causes the appearance of water to change significantly. As such, they provide a good way to test the capabilities of this method. The results were compared to a manually generated ground truth image.

The method requires setting the following parameters:

- Filter size: the size of each filter in the filter bank that is applied to the image;
- Number of textons in the dictionary ($\alpha$): the number of textons to be generated in each training area, which is the number of cluster centers of the K-means algorithm;
- Number of patches per training area ($\beta$): the number of randomly placed patches to be selected for each classification area;
- Patch size: the size, in pixels, of the training patches selected from the training areas.

The appropriate filter size and patch size are mostly dependent on the scale of the SAR image being used. For the images that were selected for the tests, the most appropriate sizes were found by performing several experiments. The patch size was 31 pixels for the filters and 71 pixels for the training patches. The other two parameters, $\alpha$ and $\beta$, were exhaustively tested to find the combination that yields the best results. Both parameters were tested using the values {5, 10, 20, 50, 100, 200, 300, 400, 500, 750}, for a total of 100 combinations for each image. For each pair, the algorithm was run ten times, using the same training areas. We found that the best results were achieved using $\alpha = 100$ and $\beta = 750$. In Table 1 we present the results obtained with these values.

The segmentation with better accuracy was obtained for Image 2 and is shown in Figure 4.

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Image</td>
<td>92.30%</td>
<td>92.72%</td>
<td>91.98%</td>
</tr>
<tr>
<td>Excl. Training Areas</td>
<td>93.27%</td>
<td>94.26%</td>
<td>92.83%</td>
</tr>
</tbody>
</table>

Table 1: Accuracy results

4 Conclusions

In this paper a new method for SAR image segmentation into water and land was implemented and tested. The method, based on textons, proved to be appropriate for the task at hand since it is based on the texture of the image instead of using pixel intensities directly, thus being more robust to the presence of noise and image artifacts due to the acquisition process.

An average accuracy across all testing images of 92.56% and a peak accuracy of 94.26% was achieved. In the future we will try to improve these numbers by testing the effects of adding DTM (Digital Terrain Model) data to aid in the segmentation process.

5 Acknowledgments

This work was funded by and developed for the FCT (Fundação para a Ciência e a Tecnologia), in the scope of the RIVERSAR project (exploration of SAR images for the improvement of floodplain models in the Tagus river), reference PTDC/CTE-GIX/099085/2008; and by project PEst-OE/EEI/LA0009/2013.

References

Mobile framework for recognition of musical characters

Rui Silva
slyzer.rui@gmail.com

Jaime S. Cardoso
Jaime.cardoso@inescporto.pt

Ana Rebelo
arebelo@inescporto.pt

Institute for Systems and Computer Engineering of Porto, University of Porto

Abstract

Programs analogous to optical character recognition systems, called optical music recognition (OMR) systems, have been under intensive development from many years. Nowadays, an interesting application of OMR concerns the online recognition, which allows an automatic conversion of text as it is written on a special digital device. This work intends to create a prototype for capturing gesture-free of musical symbols; the construction of a database collected for a set of users with different musical expertise; study of an classifier using information from the user strokes; and a comparative study between the online and the offline methods.

1 Introduction

In terms of human past knowledge, music may be one of the few things which we are certain that follow us since pre-historic times. As a way of representation, musical score notation has always been the main source of musical expression for non-hearing systems.

A musical score notation is a worldwide standard representation of music writing, grouped with a lot of a different symbols and ways to represent a very diversity of sound representations. For purposes of safeguard, digitization has been the most common tool being used to preserve a musical score, offering easy duplications, distribution and digital processing. However, a machine-readable symbolic format from the music scores is needed to facilitate operations such as search, retrieval and analysis.

OMR – Optical Music Recognition has gain a lot of contributions along the years that come since the 80’s [1], this research, mostly on offline mode, make us aware that OMR is still an important and complex field where knowledge from several field intersects.

Technology has grown exponentially along the years, and write score pieces in paper, start to be a secondary option when you have a large fan of possibilities in a computer, such as: editable scores, perfectly formatted and archived, as immediate sound response and all the other possibilities with a segmented standard music files. This led to some composers to start use computer as their main way to compose music. However, the reality is that, in an age of global technology, most of the composers still use the traditional "pen and paper" metaphor. For a complete set of 84 musical symbols, drawn 3 times, with a different thickness for each symbol, the system was showing the images that the system was showing.

An OMR system can be divided according to the input data into two categories: offline and online. One of the prominent differences is related to the type of music notation: online recognition mandatorily deals with handwritten music works. This imposes a high level of complexity in the process due to the wider variability of the objects. Besides, while an offline OMR system recognizes the symbols in a digitized music score, usually, obtained with a scanner, an online OMR system recognizes the symbols almost simultaneously when they are introduced in the system by a pen-based interface, being a real time process. The main advantages in these kinds of systems are related to: elimination of segmentation issues caused by overlapped symbols, and extra knowledge about spatial-temporal information of symbols.

This work proposes to conjugate the universal "pen-paper" metaphor with the news forms of mobile technology, as we see with the growing popularity of portable small devices, such smartphones and tablets, and their power of computing. As pen-based interfaces are in wide expansion, there is a lack of applications taking advantage of this intuitive and ergonomic way to draw musical scores, where the user composes musical scores in a traditional way by drawing the symbols on the screen.

2 State-of-Art

A method for date input in a digital system through a digital pen applied to musical software appeared at the end of the 1990’s with the Palm Pilot device [2] produced by Palm Inc. The Pilot was a small handle PDA (Personal Digital Assistant) where musical characters were inserted through a pen using a technique called Graffiti [2]. The first approach to online musical characters recognition was to create a digital pen as an interface of localization and selection, in place of the mouse, and exemplified by the work developed by [3]. Thereafter, the input data process relied on the concept of standard movements: the user needs to learn a new way to write music, not doing the same way as is used to do on paper [4, 5, 6]. Macé in [7] says that it is not possible to use a unique recognizer for all music symbols and suggest that is necessary to exploit dedicated recognizers, trying to use a mix of gestural and non-gestural symbols. Taubman in 2005 developed a system called MusicHand [8], aiming to help in the process of pitch and symbol segmentation, where the system upon entry of an unknown stroke, make the decision by comparison with a set of trained gestures, without hesitation or coaxing of the system. In 2010, Kian Chin Lee proposed a recognition process using a set of different HMMs (Hidden Markov Models) in a non-gestural approach where the users do not need to learn any special gestures for input [9].

Most of the documentation uses a unistroke (gestural approach, just one stroke) system, forcing the user to learn a new language, being an influence on the process of composition. On the existing multistroke (non-gestural approach, more than one stroke) systems, there is a lack of high level technics of image and pattern recognition, as also a set of features that were not exploited from the dynamic user handwriting.

The focus in this work will be on exploit most of the described flaws in a non-gestural approach and introduce other methods in order to improve the Online recognition process in OMR.

3 Work

3.1 User Hand-Drawn Database

The project began with a creation of a graphical user interface in android developer language, where were collected a several information of the user draws in order to create a user hand-draw database of music symbols.

The user is asked to draw symbols, with his own handwriting, according with the images that the system was showing.

For every drawn symbol, it was saved: the action done by the user (touch, drag and lift), the position X and Y, the Time and the Pressure, for every amount of milliseconds (depending on the Android machine in use). For a complete set of 84 musical symbols, drawn 3 times, with a different thickness, for every user, we can obtain a set of different handwriting. When a draw is finished, the system knows exactly where and when the path was made.

Fig. 1 - Example of a hand-draw symbol and part of the data saved from it.
We collected data from 50 users with ages between 13 and 70, from different origins and different music educations. We describe the level of knowledge in sheet music handwriting of each user, from 1 to 5, receiving handwritings from different music experiences. For purposes of segmentation and symbol analysis we included also in the database the hardware model of the machine and the resolution being used.

We define a set of 20 classes representing all of the symbols, according to their meaning, and a set of subclasses for each class.

Taking into account all of the draws done by the 50 users, we can analyse some information and predict several parameters with the purpose of restringing that symbol to those values in a future recognition process.

When facing information by a hand draw symbol, there are a lot of problems we need to face before making a real analysis. Despite being an important point to have different kind draws from different people, we cannot have control of how the user draws his symbol. From little points made intentional, to the lack of experience in touch environment or even the need to write exactly like in a paper, we have a lot of data that is not good for analyse. These “errors” must be fixed before entering the phase of segmentation, in order to have a most cohesive database.

We can see the difference in the user’s draws from the example in figure 2 from the symbol G clef.

![Fig. 2 - Difference in the symbol G clef](image)

### 3.2 Creation of the classifiers

The online nature of the input signal contains rich information about dynamics of the drawing. In order to be possible to compare two different symbols, we need to segment that information in a base of features.

We used a set of features named HBF49 [10], that are designed as a generic representation of symbols. We can extract 49 features from every symbol and use them as base of comparison. We also used a method for representing online signatures using interval-valued features [11], based on mean and standard deviation, the interval represents the upper and lower limits of a feature value. For the classification, we did not use neither the time variation nor the pressure (despite being record in the database for future use), but both could be useful to improve the recognition system.

We created a dataset using 49 features from every symbol and 1 value for the class number of the symbol, and based on the described methods we try to create 3 different neural networks to see which of them would increase the rate of recognition. In one, we concatenate the 49 features with the interval-valued method before creating the neural network, which we call ‘cNN’; naming ‘mixNN’, we use the 49 features to create the neural network and made the constrains with the interval-valued features in the last layer of the network; in last we create a simple neural network using only the 49 features from the HBF49.

In the offline process, we do not have a set of features, but a set of images. We use the OtSU’s method [12] of global threshold to convert the image to binary. The image is resized to 60X60 in order that all the symbols have the same size and be possible to compare them.

Then, we create a simple neural network with the same parameters as the one used by the online mode (using the images as input) to classify the symbols. In an offline mode, it is not possible to use the same methods, because we do not have dynamic features as in online mode.

### 3.3 Results

We achieve the best rate in online mode of 85% by cNN and the worst in mixNN with 64%. In comparison with the offline mode, the results were very superior, where we see rates in the 26% for an offline neural network.

<table>
<thead>
<tr>
<th>Performance expected in real life (%)</th>
<th>cNN</th>
<th>mixNN</th>
<th>Simple NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple NN</td>
<td>85</td>
<td>64</td>
<td>79</td>
</tr>
<tr>
<td>Simple NN</td>
<td>79</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Simple NN</td>
<td>79</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Simple NN</td>
<td>79</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Simple NN</td>
<td>79</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

In the subclasses rates we reached several in rates between 98% and 100%.

<table>
<thead>
<tr>
<th>Performance expected in real life (%)</th>
<th>cNN</th>
<th>mixNN</th>
<th>Neural Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-Class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>99</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>5</td>
<td>77</td>
<td>60</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>81</td>
<td>63</td>
<td>74</td>
</tr>
<tr>
<td>7</td>
<td>98</td>
<td>89</td>
<td>98</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>10</td>
<td>98</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>11</td>
<td>87</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>99</td>
<td>49</td>
<td>99</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>49</td>
<td>99</td>
</tr>
<tr>
<td>16</td>
<td>80</td>
<td>73</td>
<td>79</td>
</tr>
<tr>
<td>19</td>
<td>92</td>
<td>87</td>
<td>90</td>
</tr>
<tr>
<td>20</td>
<td>92</td>
<td>87</td>
<td>89</td>
</tr>
</tbody>
</table>

### 4 Future Work

With the classifiers created we intend to implement them into an android application, in order to create a software for recognition of music symbols.

### 5 Acknowledgments

This work is financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281

### References

[1] S Macé, E Anquetil, E Garrivier, B Bossis. —A Pen-Based Musical Score Editor - In proceedings ICMC, Barcelona, 2005
Abstract

Cardiovascular variability and homeostasis control in response to precise noxious stimuli and different analgesic doses were analyzed. Cardiovascular variability was altered in response to noxious stimulation, and amplitude responses varied in a dose-dependent manner with the analgesic. Responses were more pronounced to laringoscopy/intubation, when compared to tetanic and incision stimuli. Homeostasis control was also altered in response to stimulation, demonstrating that dynamic cardiovascular control relations are modified in response to precise noxious stimuli and anesthetic drugs.

1 Introduction

Pain is a multifaceted phenomenon, imperative for organism preservation. Nonetheless, this defense mechanism may be in many clinical situations undesirable, making an adequate monitoring essential. Several methods have been proposed to describe the nociception/anti-nociception balance (Noc/ANoc) including the study of heart rate variability [5], and hypnosis indexes variability [6], although none has been widely disseminated.

We propose to investigate cardiovascular variability measures [4] in response to standardized stimuli, for different analgesic doses, and assess the sympathetic/parasympathetic balance of cardiovascular dynamics.

2 Materials & Methods

Data was collected following institutional approval and written informed consent, at Hospital the Santo António (Centro Hospitalar do Porto). The study was designed to analyze responses to precise noxious stimuli, considered similar between patients (laringoscopy/intubation, tetanic stimulus, and incision), under three different analgesic doses, and the same hypnosis state, under general anaesthesia. Patients were randomly divided into three study groups considering the analgesic dose target (2, 3 and 4 mg/ml Ce).

2.1 Data Pre-Processing

Data was pre-processed to extract relevant information linked to noxious activation: from the ECG, QRS complexes and RR intervals (RRI) were detected; from the BP wave, systolic blood pressure was extracted (SBP); and finally, from the PPG, wave amplitude was extracted (PPGA).

A method with an interval depending threshold was used. This method sets the detection threshold \( n_i \) using the last detected complexes considering Equations 1 and 2 [7].

\[
\begin{align*}
    n_i &= \mu \bar{z}_{\max}, \; n = \theta_i, \; \theta_i + 1, \ldots \\
    \bar{z}_{\max} &= \bar{z}_{\max,i-1} + \alpha(z(\theta_i) - \bar{z}_{\max,i-1}), \; i \geq 1
\end{align*}
\]

where \( \theta_i \) is a detected complex, \( \bar{z}_{\max} \) is the average amplitude of the previously detected complexes and \( z(\theta_i) \) the amplitude of the signal of the most recently detected complex. \( \mu \) represents the fraction used in the threshold determination (\( \mu = 0.7 \)) and \( \alpha \) the rate at which the threshold may change (0.2).

A waiting time (silence period) \( DO \) was defined (\( \Delta \)max, 160 ms), and a D1 period during which ectopic beats may be detected with linearly descending threshold values \( g(t) \), as in Equation 3.

\[
n(t) = \begin{cases} 
    \Delta \text{max} - \theta_i - D0 - 1 & \text{if } \theta_i + 1, \ldots, \theta_i + DO \\
    \Delta \text{max} - \theta_i - D0 + 1, \ldots, \theta_i + DO + 1 & \text{if } \theta_i + D0 + 1, \ldots
\end{cases}
\]

This method was applied to the RRI detection in the collected data. Upon detecting the QRS complexes, RRI sequence was extracted. The RRI was defined at the moment of QRS detection \( t_{\text{QRS}} \) as the time difference (ms) between two consecutive QRS complexes \( t_{\text{QRS}} - t_{\text{QRS} - 1} \). A similar technique was applied to the remaining waves, extracting beat-to-beat information of the SBP (mmHg) and PPGA. Beat-to-beat PPGA was extracted considering the consecutively registered wave values at local minimum A \( t_{\text{A}} \) and local maximum B \( t_{\text{B}} \), given by Equation 4. Results shown in Fig. 1.

\[
    \text{PPGA}(t_{\text{B}}) = \text{PPGA}(t_{\text{A}}) - \text{PPGA}(t_{\text{A}})
\]

Figure 1: Pre-processing of wave files collected during the study: electrocardiogram (ECG) and QRS complexes, blood pressure wave and systolic blood pressure peaks, and photoplethysmography wave (PPG) and wave amplitude (A local minimum and B local maximum PPG points).

2.2 Cardiovascular Variability Analysis

For each stimuli defined, a period 3 minutes anterior and posterior to stimulation was analyzed. Heart rate variability measures were extracted according to [4], and composed of time based metrics. The metrics analyzed were the mean RRI values and standard deviation of all RRI (SD). Similar metrics were extracted for both SBP and PPGA. Baseline measurements were also retrieved for each patient.

2.3 Homeostasis Control

To maintain homeostasis, compensatory mechanisms regulate heart rate and vasoconstriction/dilatation to balance blood pressure fluctuations (negative feedback). To explore the homeostasis control the following metrics were analyzed: area under the curve (AUC) of RRI, SBP and PPGA [3], from baseline of each analyzed window (ARRI, ∆SBP, ∆PPGA) [1]: slope of the robust linear regression between ∆SBP and ARRI [2].

The metrics were calculated using a 30s time window and analyzed in the periods prior and following incision, since in the majority of the cases, invasive blood pressure measurements were only available following induction and laringoscopy/intubation.
Table 1: Median cardiovascular variability amplitude variations for baseline, anterior and posterior stimuli periods, for each study group (remifentanil effect-site concentration of 2.0, 3.0 and 4.0 ng/ml; n=10,11,10): baseline, and pre-laringoscopy; pre-laringoscopy and post-intubation; pre-tetanic and post-tetanic; and pre and post-incision. (NA - not applicable)

<table>
<thead>
<tr>
<th>Group</th>
<th>RRI Mean SD</th>
<th>Laring/Int Mean SD</th>
<th>Tetanic Mean SD</th>
<th>Incision Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>130.72 -14.70</td>
<td>118.83 -13.95</td>
<td>45.54 -2.68</td>
<td>23.47 -3.10</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBP</td>
<td>13.89 -9.54</td>
<td>26.20 -8.5</td>
<td>49.13 -4.3</td>
<td>9.87 -0.99</td>
</tr>
<tr>
<td>2</td>
<td>1.37 -2.49</td>
<td>8.81 -0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPGA</td>
<td>-0.46 -0.32</td>
<td>0.37 -0.06</td>
<td>-0.34 -0.05</td>
<td>0.01 -0.03</td>
</tr>
<tr>
<td>3</td>
<td>0.01 -0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 Statistical Analysis

Non-parametric paired sample tests were used for the comparison of the values before and following stimulation, since measurements inside the same patient were compared (paired sample median test, Wilcoxon signed ranks). For the assessment of the analgesic impact, a non-parametric median test was employed, since in this case comparison between different groups of patients is performed (Kruskal-Wallis and Mann-Whitney median tests). P<0.05 was considered significant.

3 Results and Discussion

Data from 34 patients were collected following institutional approval and written informed consent. No difference found between groups considering average age, weight, height and gender distribution.

3.1 Cardiovascular Variability

Statistical differences were found between RRI anterior to laringoscopy and posterior to intubation, for all study groups, with an increase of heart rate post-stimulation. No statistical difference was observed for the tetanic (increasing tendency) and incision stimuli. Also a statistical significant increase in RRI values at induction was observed in all study groups, due to the introduction of anesthetic drugs, without the input of noxious stimuli, leading to a generalized depression. This response was observed in all cardiovascular variables (decrease in SBP and increase in PPGA).

In response to laringoscopy/intubation and incision SBP values increased, while for the tetanic stimulus a tendency of decreasing SBP was observed with the increasing doses of analgesic.

PPGA was decreased in response to the laringoscopy/intubation and incision stimuli. Although no statistically significant responses to tetanic stimulus were observed, it may be observed that in remifentanil group 1 a small tendency towards PPGA decrease is observed, nevertheless this tendency was inverted with the increase of remifentanil doses.

Statistical significant differences were found for RRI amplitude response to stimulation, being laringoscopy/intubation the stimuli with most pronounced response, followed by incision (Table 1).

It is important to notice that baseline values varied between patients, also the baseline following drugs administration, which may difficult the definition of an adequate state of the Noc/ANoc balance.

3.2 Homeostasis Control

The homeostasis control was assessed by the methods described earlier, using RRI, SBP and PPGA series. Table 2 shows the obtained results (AUC and regression slope) for each study group, in response to incision (difference between post and pre-stimulus periods).

AUC of ARRI and ∆PPGA were decreased in response to stimulation, and the ∆SBP AUC was increased. The slope of the robust regression between ∆SBP and ARRI was decreased following incision, demonstrating that cardiovascular regulation mechanisms are altered following stimuli.

Fig. 2 shows a trend of the slope of ∆SBP and ARRI regression, with overlapping stimuli onset. Pronounced decreases in the estimated slope, following incision, are observed.

Table 2: Homeostasis assessment in response to the incision stimulus (median variation between posterior and anterior periods), for each study group (remifentanil effect-site concentration of 2.0, 3.0 and 4.0 ng/ml).

<table>
<thead>
<tr>
<th>Group</th>
<th>AUC(RRI)</th>
<th>AUC(SBP)</th>
<th>AUC(PPGA)</th>
<th>Slope(SBP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.85</td>
<td>-0.93</td>
<td>-1.87</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.68</td>
<td>-0.11</td>
<td>-2.69</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-1.84</td>
<td>0.99</td>
<td>-1.37</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: ∆SBP and ARRI estimated robust regression slope in a patient of the study. Full triangles present times of stimulation for laringoscopy, intubation, tetanic and incision.

4 Conclusion

Cardiovascular variability during general anesthesia, for different precise noxious stimuli, under three different analgesic doses was assessed. The physiological signals analyzed responded both to stimulus, and analgesic attenuation, in a dose dependent manner, which is the basis of a tool to monitor the Noc/ANoc balance.

Although the stimulus onset can not be predicted, increased variability and altered cardiovascular control may be used to predict movement [1], as well as parasympathetic/sympathetic imbalance [3]. Analysis of these events needs to be addressed in future studies, under steady-state conditions.

ACKNOWLEDGMENT

Work partially funded by the project I-CITY - ICT for Future Health NORTE-07-0124-FEDER-000064.

References


Quality measures for iris images in mobile applications

Ana F. Sequeira¹
afps@inescporto.pt
Juliano Murari²
juliano.murari@usp.br
Jaime S. Cardoso¹
jaime.cardoso@inescporto.pt

¹ INESC TEC and Faculdade de Engenharia Universidade do Porto
Porto, Portugal
² Universidade de São Paulo

Abstract

Biometric recognition is nowadays a mature technology with several applications. However, during the past few years, biometric quality measurement has become an important concern. Concerning iris recognition, there are many factors which may affect the quality of the iris images. One of the main challenges is the performance degradation in less controlled situations. The proliferation of portable hand-held devices with at-a-distance and on-the-move biometric acquisition capabilities lead to the necessity of assessing the quality of the images prior to the recognition process. In this work we applied state-of-the-art methods to compare three quality measures obtained for iris images from two databases: a widely used database and a recently built one constructed with a hand-held device.

1 Introduction

Nowadays, more than ever, the recognition of persons made in an automatic, yet reliable, way is an attractive goal. In most of our daily activities personal identification plays an important role. The most traditional techniques for personal identification rely on two kinds of items: a personal item, like a passport, driver’s license, ID card, credit card or a simple set of keys, or a personal knowledge that, theoretically, are not accessible to others, such as passwords or personal identification numbers. These approaches present the obvious disadvantage that these items may be lost, stolen, forgotten or misplaced and passwords can easily be forgotten or accessed by an unauthorized user. Biometrics represents a return to a more natural way of identification: many physiological or behavioral characteristics are unique between different persons. Testing someone by what this someone is, instead of relying on something he owns or knows seems likely to be the way forward [8].

Biometric recognition is a mature technology used in many government and civilian applications such as e-passports, ID cards, and border control. Examples include the US-Visit (United States Visitor and Immigrant Status Indicator Technology) fingerprint system, the Privium iris system at Schiphol airport, and the SmartGate face system at Sydney Airport [2]. However, during the past few years, biometric quality measurement has become an important concern. Studies and benchmarks have shown that biometric signals’ quality heavily affects biometric system performance [2]. However, when compared with other steps like feature extraction and pattern recognition tasks, this important step has nevertheless received little attention. In fact, several factors can affect biometric signals’ quality, and quality measures can play many roles in biometric systems. One of the main challenges facing biometric technologies is performance degradation that occurs when the data is collected in less controlled situations. We observe nowadays a proliferation of portable hand-held devices equipped with at-a-distance and on-the-move biometric acquisition capabilities. These conditions are just two examples of non-ideal scenarios in which biometric recognition is not sufficiently mature. These will require robust recognition algorithms that can handle a range of changing characteristics. [2]

Iris recognition is a particular type of biometric system that can be used reliably in identifying a person by analyzing the unique patterns found in its iris. [3] The pioneer works in iris recognition set the basis of the typical iris recognition system architecture: segmentation, normalization, feature extraction and matching [5]. For an overview of the system see Fig. 1.

There are many factors which may affect the quality of the iris images. Iris images can be affected from a wide range of qualities like dilation, specular reflection, iris resolution, motion blur, camera diffusion, presence of eyelids and eyelashes, head rotation, camera angle, contrast, resolution or others. Despite the less attention this aspect has received compared to other aspects of iris recognition, there can be found in literature some works on iris quality assessment made by researchers like John Daugman, Hugo Proenca, J. Zuo, N. Kalka and N. Schmid [3].

From a biometric point of view, the quality of iris images can be assessed by measuring one of the following properties: i) focus, ii) motion blur, iii) occlusion, and iv) others including the contrast or the dilation of the pupil. A number of sources of information are used to measure these properties such as the high frequency power spectrum, angle information provided by directional filters, pixel intensity of certain eye regions, or different ratios comparing the iris area to that of the image, or the iris and pupil sizes. Iris quality can be assessed either analyzing the image in a holistic manner, or combining the quality from local blocks of the image. [6]

2 Implemented Methods

Some of the measures are obtained from the entire eye image but others are extracted only from the iris region, therefore a segmentation step is required. We choose to make the segmentation process manually, in order to ensure reasonable accuracy. The manual segmentation is done by marking three different points in the image. The first point is the eye center, i.e., we consider a single center for both pupil and iris (which is a somehow abusive assumption since the iris and the pupil are not necessarily concentric). The second point is marked in the pupil border and the third in the iris border. With these points is possible to determine the iris and pupil radius and then approximate the contours as two concentric circles. With the manual segmentation’s information we are able to map the regions of interest which will be eventually used by the algorithms.

2.1 Algorithm 1 - High Frequency Power

The High Frequency Power algorithm, which provides quality measure 1, works on the whole image and measures the energy concentration in the high frequency components of the spectrum using a high pass convolution kernel of 8x8. The application of this convolution is a good Fourier Transform approximation and works as high frequency spectral analysis, which can be considered an estimator of focus [4].

The final measure is given by equation (1), where $I_i$ is the convolution matrix $M \times N$.

$$ r = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} I_i(i,j) $$

2.2 Algorithm 2 - Local Contrast

The Local Contrast algorithm, which provides quality measure 2, is based on the bounding box that involves the iris and the pupil. This bounding box

Figure 1: Iris recognition scheme, adapted from [9]
box is divided in blocks of $P \times P$ and for each block it is applied the Fast Fourier Transform (FFT) algorithm to extract the medium power frequencies, which better represents the contrast. The final value is given by the number of blocks with medium values (between 20 and 60) divided by the total number of blocks. This algorithm was inspired in an occlusion estimation technique [1] and it was adapted for contrast estimation in [6] where more details can be found.

2.3 Algorithm 3 - Global Contrast

The Global Contrast algorithm, which provides quality measure 3, explores the fact that parts extremely bright or dark of the image are not useful and can be considered as noise. Thus, pixels near medium value (128 in 8-bit image) are considered of best contrast [1]. In order to quantify the contrast, the original pixels values are normalized between 0 and 25 (Figure 2). Original pixels near medium value will get higher values in the normalized scale, as well as very low and very high values (< 10 and > 245) are normalized to 0. This measure was presented in [1] and it was adapted for global contrast estimation in [6] where more details can be found.

Figure 2: Normalization function of the algorithm 3.

3 Experimental Setup

3.1 Datasets

The implemented methods were tested in two datasets. We used 450 images from the UBIRIS.v2 database [10]. This database was created with the major purpose of constituting a new tool to evaluate the feasibility of visible wavelength iris recognition under far from ideal imaging conditions. The images from UBIRIS.v2 were captured at-a-distance and on-the-move and therefore they contain several kinds of noise. The other dataset is a subset of 800 iris images from a newly built database, the MobBIO Multimodal Database [7]. The MobBIO has the particularity of being a database acquired with a hand-held device. The equipment used for the samples acquisition was an Asus Transformer Pad TF 300T. The RGB images were obtained with a 8 megapixel resolution camera with auto focus. The iris images were captured in two different lighting conditions, in a room with both natural and artificial sources of light, with variable eye orientations and occlusion levels, so as to comprise a larger variability of unconstrained scenarios.

3.2 Experimental Results and Discussion

The algorithms implemented return a set of 3 quality measures calculated for each database. In Table 1 we present the mean value ($\mu$) and the standard deviation ($\sigma$) of each of the three quality measures for each of the two databases.

<table>
<thead>
<tr>
<th>Database</th>
<th>Alg.1</th>
<th>Alg.2</th>
<th>Alg.3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$\sigma$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>UBIRIS.v2</td>
<td>67.63</td>
<td>5.98</td>
<td>0.71</td>
</tr>
<tr>
<td>MobBIO</td>
<td>75.47</td>
<td>7.51</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The range of values of each quality measure differs in the two databases tested even though they have the same order of magnitude. The standard deviation of the values ($\sigma$ in the table) is higher for the MobBIO database. This fact is somewhat coherent with the less constrained acquisition conditions in the MobBIO construction when compared to the UBIRIS.v2.

Also, the equipment used has more factors varying such as camera focus and the acquisition in this scenario is much more influenced by the $modus operandi$ of the user. The more significant difference appears to be found in the quality measure 2, “local contrast”. This may be due to the type of camera used to capture MobBIO’s images which causes less sharpness in the images. The MobBIO database has a better result than UBIRIS.v2 for quality measure 1 but for the others the values are worse. More research is required before we can hand a clear and complete interpretation of the results obtained.

4 CONCLUSIONS AND FUTURE WORK

The actuality of the iris images quality assessment topic is unquestionable. As the field of application of iris recognition broadens, to embrace the demands of a society highly dependent on mobile and portable devices, the necessity of improving the reliability of the recognition process arises. Under these challenging conditions (recognition made by portable devices such as mobile phones or web cam and in indoor or outdoor scenarios) the quality of biometric samples, iris images in particular, has a crucial importance to the success of the recognition process. The existing databases offer, at a satisfactory level, the quality necessary for the recognition, however, this is not guaranteed when we consider the images captured in mobile applications for instance.

This work was a preliminary approach to the problem of quality assessment of iris images acquired with a hand-held device. We used a new database with iris images acquired in unconstrained conditions and with a hand-held device. This database was tested for three state-of-the-art methods in image quality assessment and the results were compared with the ones obtained for an existing and widely tested database for iris recognition. As expected the range of values obtained for each quality measure differs in the two databases tested.

We intend in future works to enlarge the set of quality measures to comprise different state-of-the-art methods. Another path we aim to follow is the quality assessment of a quality measure itself, which we may accomplish by indirect evaluation using the results (ideal versus real) of a recognition process.

Acknowledgments

This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281 and PhD grant with reference SFRH/BD/74263/2010. The second author would like to thank the National Council for Scientific and Technological Development (CNPq) - Brazil.

References

Abstract

In this article we suggest a content-based image retrieval (CBIR) approach to automate the identification of meals and caloric estimation. We analyze the performance of the local features descriptors Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). To complement the SIFT and SURF's lack of colour information in the description of the key points, we propose the description of the key points in each colour layer, the concatenation of a colour histogram in each key point descriptor and the use of global colour histograms. To handle the local descriptors, we analyze a Bag of Words (BoW) approach, and we propose the use of BoW by class to take advantage of the intrinsic division of the meals' classes. To estimate the calories, we propose the use of the number of local descriptors as a comparison factor. Overall, the standard methodologies offer an interesting foundation for meals recognition. Adding the suggested methodologies, namely the introduction of colour information and application of BoW by class, contributes to a significant increase in performance. In terms of calorie estimation, the proposed methodology can be satisfactory for simple scenarios.

1 Introduction

It is noticeable an increasing investment in health and well-being by the society in general, for instance, following healthy diets. However, it is not always easy to keep a cared nutrition and much less a controlled one. By using Computer Vision, we intend to introduce a model to assist the maintenance of a more controlled diet. The objectives of this project consist in adapting the state-of-the-art computer vision algorithms for content-based image retrieval (CBIR) to identify meals and define statistical methods to estimate the caloric value of meals. We expect with this project to provide a model to automate the identification and caloric estimation of meals based on images that is flexible enough not to require any type of control in the way the images are collected and not limit the set of meals identifiable.

2 Related Work

Some of the interesting proposals for meals identification and caloric estimation include Shroff's automatic way to recognize food [1]. In her work, the background is removed using adaptive threshold and then the connected components are classified using 4-Means. The mask obtained is dilated and eroded to remove unwanted noise and then applied to the original image to separate the background from the objects. Features based on colour, size and texture are extracted from the objects, and used for classification with Neural Networks. The requirement of keep a space between foods and ensure that the background has a uniform colour, lighter than any food present in the scene, removes flexibility to the model. Sun proposed a method to determine the portion size of food [2]. In his proposal, the user identifies and measures the food upper surface, length and height, which is followed by segmentation with active contours. Then, the camera is calibrated to measure the segmented objects, which implies determining the camera’s focal length, which is estimated from the vanishing points. The amount of food is calculated by taking the product of the estimated area and height. The model produces results with considerable precision, however the estimate is not completely automatic since an initial interaction with the user is necessary.

3 Methodologies

In CBIR systems, images are indexed by the visual content, which is extracted using local/global feature descriptor algorithms. The use of global and local features in parallel allows for a more solid and reliable characterization of the image. The traditional process of research consists in given a query image, calculate it's similarity to the images stored in the database and select the most similar.

3.1 Detection and Description of Local Features

A local feature in an image is a pattern that differs from its immediate neighbourhood, usually associated with a change of one or more properties simultaneously, such as intensity, colour and texture. The process of feature extraction has to determine not only the location of the key point, but also the size and shape likely. Of all the properties required of a local feature descriptor algorithm, the most important is undoubtedly repeatability. Two of the most standard local feature descriptor algorithms are SIFT and SURF.

3.1.1 SIFT: Scale-Invariant Feature Transform

In SIFT the detection of the key points is based on calculations of difference-of-Gaussians. For the description, the algorithm uses oriented gradient histograms to characterize the local neighbourhood of these key points [3].

3.1.2 SURF: Speeded Up Robust Features

In SURF the location of the key points is determined through the use of Hessian detector. This algorithm calculates the orientation using Haar Wavelets for each key point. SURF forms the descriptors from pulses of the points around the key points [4].

3.2 Detection and Description of Global Features

A global descriptor characterizes globally visual features such as colour, texture or form. One standard technique used to describe colour consists in the use of a colour histogram [5], which requires a process of quantization and therefore a reduction of detail. In the construction of a colour histogram, for each pixel, the column (in the histogram) with index equal to the pixel colour quantized is incremented in one unit.

3.3 BoW: Bag of Words

In the detection of local features we obtain a large number of descriptors, so a comparison descriptor to descriptor is not feasible especially for a large database (DB) of images. One alternative is the use of Bag of Words [6]. This method selects N reference descriptors (words) used as a basis to construct a histogram of N bins for each image. Each word corresponds to a bin in the histogram. For an image, each of its descriptors is compared with each word, incrementing the bin associated with the word that is more similar to the descriptor. With this method, the image search is done by comparing pairs of histograms rather than comparing pairs of descriptors.

4 Algorithms

Due to the physical features of the meals, colour information can contribute to improve categorization, however both SIFT and SURF perform the detection and description of the key points on a gray scale version of the original image, discarding colour information. Another factor that can help in the categorization is the intrinsic division between classes of meals. This is lost with the use of BoW once all descriptors are mixed during the sampling process. Regarding the calories, we can use the information of the key points as basis to perform estimation.

4.1 Introduction of Colour Information Locally

- Description of the key points in each Colour Layer (DCL); After detecting the key points, the image is divided in the RGB layers and the position of the key points mapped for the homologous location in each layer. Then, for each key point, we make their description in each colour layer, and concatenate the result in a single descriptor.
- Concatenation of Colour Histogram (CCH); Once determined the location and size of the key points, we construct for each key point a colour histogram of the area around the point, and then concatenate the histogram with the base descriptor.
4.2 Bag of Words by Class

Once the meals are grouped by class in the DB, it is possible to avoid this situation by constructing one Bag of Words for each class of meal (for example, one BoW for Pasta Bolognese, another for Steak with Chips, and so on). To classify an image, we build a histogram to the (query) image, with as many columns as the number of existing classes. For each descriptor of the image, we determine the BoW that contains the word most similar to the descriptor, and then we increment the column of the histogram that corresponds to the BoW's class. The picks in the histogram are the classes with higher probability of match the class of the query image.

4.3 Calorie Estimation

The estimation of the calories can be done by comparing the number of local descriptors extracted from the query image with the average number of descriptors extracted from the images on the DB belonging to the same class of the query image. As an example, assuming there is an image of a meal with a portion equivalent to 646 calories. If we extract descriptors from this image, then, in another image of the same meal type (class) but with twice the amount of food, we expect to extract twice the number of descriptors. This approach allows obtaining calorie estimation for simple scenarios. In more complex scenarios in which there is occlusion of food or where the ratio between foods varies precise estimates are not expected.

5 Application

For each image, we perform the detection and description of the local features using SIFT and SURF. We analyze the advantages of introducing colour information globally with a colour histogram (GCH) covering the RGB (R,G,B)\(^3\) and HSI (H,S,L)\(^1\) models, and locally with DCL and CCH. In CCH, for each key point we construct a colour histogram contemplating the RGB (R,G,B) and HSI (H,S,L) models. We test the performance of the standard BoW with 50, 100 and 200 words and BoW with 50 and 100 words per class. A K-Means algorithm is used to perform the selection of the words. We give a weight of 80% to the local features and a weight of 20% to the global features. The weights were set based on the relevance of each histogram.

6 Experimental Results

For the meals' identification, we used 17 sets of images, each set with five images from one specific class of meal. In each set, we selected one image randomly to serve as query image, and used the remaining to build the DB. For the calorie estimation, we used 17 pairs of images, each pair of one specific meal class and with similar amounts of food. For each pair, we compared the difference between the number of descriptors extracted. We collected the images using a 2MP phone camera (LG GS290). The photos were taken without any kind of special control over the camera position, lighting, etc. We can see the results of the meal identification in Table 1, and calorie estimation in Table 2.

<table>
<thead>
<tr>
<th>SIFT</th>
<th>Average</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50 Words</td>
<td>71.6 %</td>
<td>5.67 %</td>
</tr>
<tr>
<td>SURF</td>
<td>82.1 %</td>
<td>5.57 %</td>
</tr>
</tbody>
</table>

Table 2: Average and standard deviation of the difference on the number of descriptors between each pair of images, with SIFT and SURF.

In identification, SURF ensured a better distinction between classes than SIFT. The introduction of colour information globally with GCH and locally with DCL and CCH improved the discrimination between descriptors. Of the settings based in BoW, the use of 50 words per class produced the best results. With the best settings, we achieved a precision globally higher than 80%. Regarding the estimation of calories, we obtained an average deviation lower than 10% with SIFT and SURF.

7 Conclusions

SIFT and SURF provide a solid description of local features in images. Both our proposals to compensate the lack of colour information in the description of the key points (DCL and CCH) ensure a better performance in distinguishing between descriptors, especially CCH. The GCH helps to increase the accuracy in meals recognition. The HSI model provides better overall colour description than the RGB model. In handling local descriptors, an approach of Bag of Words per Class is superior to the standard approach, as it maintains the distinction between classes of meals. In both approaches, using a reduced core of words led to better results, possibly because using fewer words increased the distinction between the words, leading to greater discrimination. In terms of calorie estimation, the proposed methodology can be satisfactory for simple scenarios. However, in more complex situations it is expected that the estimated values deviate much from the real caloric value.

Acknowledgements

This work was partially supported by the project PEst-OE/EEI/LA0008/ 2013 from Instituto de Telecomunicações.

References

Face Recognition with Neural Networks Classifier using SIFT and SURF Descriptors

João Sargo, João. R. Caldas Pinto, João M. Costa Sousa
joao.sargo@hotmail.com, jcporto@dem.ist.utl.pt,
jmsousa@ist.utl.pt,jmsousa

Abstract

Face recognition has many important practical applications, like surveillance and access control. In this paper we propose an algorithm using SIFT (Scale-Invariant Feature Transform) and SURF (Speeded Up Robust Features) descriptors for the recognition process with neural networks classifiers, with very promising results.

1 Introduction

Machine recognition of human face from still and video images has become an active research area over the last 40 years [1] in the communities of image processing and pattern recognition. This interest is motivated by wide applications, including mug shot matching, surveillance, access control, personal identification, and forensic/law enforcement applications [2]. In this paper, given a database of face images from a group of people and given an unknown face image, we want to answer the question: “which person in our database does this image belong to?” Many algorithms and techniques have been proposed for solving such a problem [2].

2 Databases

In this work we used two databases: the benchmark AT&T face database [3] and the IST database of faces created for this work. The AT&T database was created between April 1992 and April 1994 in the laboratory of the University of Cambridge. It contains 400 images of 40 persons with 10 images/person. There are different orientations and facial expressions for each subject as shown in the Figure 1.

The IST database was created using a kinect in Technical University of Lisbon, Instituto Superior Técnico, in November 2012. It contains 100 images of 10 students with 10 images/person. The use of the kinect for this task allowed the record of the 100 images in gray scale and in depth.

3 Methods

The most popular face recognition approaches such as eigenfaces [4] and fisherfaces [5] make use of appearance based projection methods, like PCA (principal component analysis), ICA (independent component analysis) or LDA (linear discriminant analysis). In the present work, we used descriptors extraction algorithms; these descriptors after treatment were then used as inputs to the classification neural network (NN) models.

We used four different descriptors extraction algorithms: the first one based on the SIFT algorithm created by Lowe [6], the second based on the SURF algorithm, first presented by Herbert Bay et al. in 2006 [7], and the others correspond to two variations of the SIFT and SURF algorithm as in T. Coito [8]. For a given image, these algorithms detect the points of interest and then extract the descriptors for each point. While the points of interest vary from image to image the descriptors are constant depending on the algorithm. Although neural networks [9] do not uncover internal relations of physical variables, they are robust to the presence of noisy data and have good classification performance.

4 Approach

As described before, for each image the number of points of interest detected is variable. However, the dimension of the inputs of the NN models has to be constant. Thus we had to ensure that for each image the number of evaluated descriptors is the same. To do so, let us assume there were found n interest points characterized by m features each (usually m= 64 or 128). With this data we build a 2m feature vectors whose elements are evaluated through equations (1) and (2). For each feature j (j=1, 2, ..., m), the arithmetic mean μj (eq. 1) is computed. Where \( a_{ij} \) corresponds to the value of the point of interest i, considering the feature j.

\[
\mu_j = \frac{1}{n} \sum_{i=1}^{n} a_{ij}
\]  

(1)

The standard deviation \( \sigma \) was then computed (eq. 2).
\[ \sigma_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (a_{ij} - \mu_j)^2} \] (2)

Thus, for these equations (\( \mu_j \) and \( \sigma_j \)) result a 128 or 256 constant dimension input, regardless the number of points of interest. In the case of the database with grey scale and depth images the process was the same, but computing the mean and standard deviation for the grey scale images and for the depth images separately, resulting a 256 or 512 constant dimension input.

For each database considered, 60% of the samples (images) were chosen as train set, 20% as validation set and 20% as test set. This division was random and made in a way that each class (person) had the same percentage of images in each set. It was decided to use one hidden layer on the NN models, these models operate as universal approximators. After training, for a given image, the NN models returned a value between 0 and 1, considered the matched person to that image. The class with the higher matching value was the matched person to that image.

5 Results

Even though we had presented two databases of faces, we used the methods and the approach described before in three different sets of images: the grey scale AT&T database, the grey scale IST database and the grey scale and depth images of IST database. For each database, a study was performed to select the number of neurons for the hidden layer, i.e. lower amount of neurons without losing considerably accuracy in the results of correct matches on the test data, lowering the level complexity of the models.

After the best set of parameters was selected for each model, 100 rounds of the process were performed, always with different partitions in the training, validation and test sets. The model with better performance (accuracy) was selected. The following results show the accuracy of correct matches for the training set and some additional information about the models created.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy [%]</th>
<th>Matches:</th>
<th>Neurons:</th>
<th>Dimension of the input</th>
</tr>
</thead>
<tbody>
<tr>
<td>openSURF</td>
<td>97.5</td>
<td>18 out of 80</td>
<td>40</td>
<td>128</td>
</tr>
<tr>
<td>SIFT_andrea_vedaldi</td>
<td>80</td>
<td>17 out of 20</td>
<td>20</td>
<td>256</td>
</tr>
<tr>
<td>SIFT_tiamo_coito</td>
<td>80</td>
<td>17 out of 20</td>
<td>20</td>
<td>256</td>
</tr>
<tr>
<td>SURF_tiamo_coito</td>
<td>80</td>
<td>17 out of 20</td>
<td>20</td>
<td>256</td>
</tr>
</tbody>
</table>

Table 1: Comparison of descriptors extraction algorithms for face recognition problems using neural networks classification models.

6 Conclusions

The best results were always obtained using the algorithm for extraction of descriptors: openSURF, i.e for the methodology used in this work, using the mean and standard deviation of the descriptors as inputs for the NN models, the algorithm that achieves the best relationship with these entries and the considered classes is the algorithm openSURF.

Although the results of the IST database were perfects, it should be noted that due to the size of samples (100 images vs 400 of the AT&T) the results of this database are not so conclusive as the well known AT&T database results. Comparatively with the accuracy achieved with traditional approaches, using eigenfaces or fisherfaces in the AT&T database, the performance of our models using NN were slightly better (97.5% with our models and 96.3% obtained in [10]).

We finally conclude that the use of depth images did not improved the results at least in this preliminary study.

7 Future Work

In order to make the results about the use of the depth descriptors as an improvement to the face recognition problem, the IST database should be updated with more images in order to make the results more conclusive about the use of the depth in face recognition.

It’s important to note that the methodology used is very versatile and can be applied in the most various computer vision applications, just by having a problem to classify and a considerable number of images to train the network. Thus, in the future, it is possible to use this approach with other applications.

Acknowledgements

This work was supported by FCT, through IDMEC, under LAETA.

References

Automatic Visual Inspection of Ceramic Plates based on SIFT and SURF Descriptors

João R. Caldas Pinto, Rafael Baeta, Mariana Pereira, Ricardo Laranjeira, João Sargo and Carlos Cardéira
jc.pinto@dei.ist.utl.pt; (rafael.baeta, mariana.g.pereira, ricardo.m.laranjeira@ist.utl.pt; joao.sargo@hotmail.com, cardeira@dei.ist.utl.pt

Abstract
This paper concerns the problem of automatic visual inspection of ceramic plates. It will be studied two classes of plates and three groups of defects. Based on Scale Invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF) descriptors and a simple decision system based on a neural network we got very promising results with real experiments. Due to the generality of these algorithms we expect to easily solve another type of defects in different classes of plates.

1 Introduction
A large number of pottery factories rely on human workers to most of the inspection operations that should be carried out in a production line. At first glance it seems that would be easy to automatize these inspections based on computer vision. However, due to the large variety of plates and other dishware that circulates on conveyors in different stages of the production it is almost impossible to guarantee a pre-defined stable part positioning and stable lighting conditions. It is probably due to this fact that literature about this subject is not abundant [1]. There are quite a good number of papers in ceramic tiles e.g. [2], [3], [4] deals with the inspection of different type of objects and use SIFT descriptors.

In this work we selected three groups of defects in two types of plates (Fig. 1). We define:
- A spoiled ceramic decals in one type of plates (Fig 1a)
- B missing glass on the back in the other type of plates (back) (Fig. 1b)
- C granules in the same type of plates (front) (Fig. 1c)

This paper is organized as follows. Section 2 briefly describes the main techniques used in this work. Section 3 discusses the lighting conditions and presents the derived algorithms and obtained results for the different defects. Finally, Section 4 presents the conclusions and future work.

2 Related Techniques

Scale Invariant Feature Transform (SIFT).
In a very brief description this algorithm, developed by David G. Lowe [5], [6], can be divided in the following two majors steps:
• Interest points localization.
• Build interest points descriptors.

Both steps are quite complex to achieve the claimed invariance. The first one includes building a scale space at different image sizes (octaves). This space will allow replacing the evaluation of second order derivatives by Difference of Gaussians (DOG), from where interest points with sub-pixel accuracy are evaluated. The second step involves assigning an orientation to the interest points and building their descriptors. These are evaluated collecting gradient directions and magnitudes around each one. This is carried out in a 16 by 16 window around interest points. This window is still divided in sixteen 4 by 4 windows where gradient magnitudes and orientations are put into a 8 bin histogram. Hence, this gives rise to 128 descriptors for each interest point. These descriptors still suffer some kind of normalization [6].

Speeded Up Robust Features (SURF)
SURF was first presented by Herbert Bay et al. in 2006 [7]. It is partly inspired by the SIFT descriptor. Like the SIFT method the first two steps rely on a scale-space representation and first and second order differential operators.

The originality of the SURF method is that these operations are speeded-up by the use of an integral image and box filters techniques. The interest points evaluation is based on the computation of the discrete Hessian operator at several scales using box-filters. They are these box filters that approximate the second order derivatives. For each key point, in order to achieve rotation invariance, is evaluated the dominant orientation by considering the local gradient orientation distribution, estimated with Haar wavelets. The descriptors, a 16x4 vector, are built corresponding to a local histogram of the Haar wavelet responses. Like the SIFT descriptors these features still suffer some kind of normalization.

Hough transform
Due to the characteristic shape of the plates (circular) it is quite useful the Hough transform [8] to detect circles in the images. This well-known transform was first concerned with the identification of lines in an image, but later the Hough transform has been extended to identifying positions of arbitrary shapes, most commonly circles or ellipses. In our case the problem is facilitated because we know how many circles we expect in our images and even approximately their ratio.

3 Implemented methods

3.1 Image acquisition. Illumination
The illumination of quite reflective ceramic plates is very challenging. Indeed with normal illumination the reflexes invalidate any tentative to find sometimes very small defects. Inspired in commercial solutions a low cost dome was designed and quite satisfactory images (Fig. 2) were obtained.

3.2 Defects of type A
To study the problem, a set of 24 images were used corresponding to 4 plates each one in six different positions. All the plates have defects and the goal was to find them automatically.

It was decided to base our approach classifying the points of interest found by the SIFT algorithm. After some attempts it was found that was better to apply local fine tuning to the different regions of the image, namely, the black sheet, the white sheet and the remaining plate (Fig. 4). To find these regions the images were rotate to a standard position. Notice that images in two different plates are not absolutely equal, otherwise a simple subtraction would be enough to find the defect. An examples of interest points obtained with this plate is given also in Fig 3.
The classification of the interest points obtained by the SIFT algorithm as defect or no defect was carried out by a neural network. Applying this algorithm to all the images it was achieved quite encouraging results.

### 3.3 Defects of type B

These defects are quite hard to detect. Special light conditions would help but we wish to make it as simple as possible. Because these defects correspond indeed to a given disruption to an usual pattern we tried again to use one of the SIFT inspired features. Despite using again neural networks as final classifier we follow a different approach to this type of defects. The algorithm can be described in the following steps:

- Find the region of the interest (ROI) of the plates using the Hough transform (Fig 4a).
- Find the points of interest using one of the methods described in Section 2. Select those that are inside the interest region, Fig 4b.

Let it be N the total number of points of interest of the plate k each one with m features (m=128 (SIFT) or m=64 (SURF) and define $d_k(i,j)$ the descriptor $i (= 1 \ldots m)$ of the interest point $j$.

Define a global descriptor for the plate k as:

$$g_k(j) = \frac{1}{N} \sum_{i=1}^{N} d_k(i,j)$$

Each plate is now characterized by a vector of m features. To train the neural network were used 14 images without defects and 20 with defects and the percentage allocated for training and validation was respectively 50% and 30%. The neural network is type feed forward with one hidden layer with 70 neurons.

### 3.4 Defects of type C

This type of defects corresponding to very small black spots is important to be detected at least in the central region of the plate without any decals. The solution would be trivial with ideal lighting conditions, but no classical methods resulted with real images. Again it was tried to use interest points found for one of the methods referred above. It was used 10 plates, 6 without defect and 4 with defect. This number of plates should be bigger, however, it was verified that were found interested points using the SURF method inside the ROI, only in the case of existing defect. These results (Fig. 5) are quite promising, even if you need to introduce a more powerful decision rule to deal with a bigger number of plates.

### Acknowledgements

This work, being part of the Project Produtech PSI (n.º 13849) – New Products and Services for the Manufacturing Industry, has been partly funded by the Incentive System for Technology Research and Development in Companies (SI 1&D), under the Competitive Factors Thematic Operational Programme, of the Portuguese National Strategic Reference Framework, and EU’s European Regional Development Fund. The authors also wish to acknowledge the gratitude and appreciation to all the PPS3-Produtech PSI project partners for their contribution during the development of various ideas and concepts, presented in this paper. This work was also partially funded by the Project PEst-OE/EME/LA0022/2011, LAETA Unit (IDMEC-Pole IST) of the Foundation for Science and Technology (Fundaçao para a Ciência e Tecnologia).

### References

1. Resumo

O objetivo deste trabalho consiste na análise de imagens de uma fachada de edifício, captadas por uma câmera fotográfica estática, e classificação do posicionamento dos estores de cada uma das suas janelas. De forma a tornar esta classificação automática, foi desenvolvido um programa no software Labview. Este trabalho surge no contexto da unidade curricular de Visão Computacional do curso de Engenharia Mecânica, do ramo de sistemas e de um trabalho proposto durante a minha estadia no Laboratório Nacional de Engenharia Civil (LNEC). Este trabalho integra-se num estudo, onde se pretende avaliar o comportamento humano na abertura manual dos estores, tendo em consideração as condições ambientais exteriores. Desta forma será possível no futuro automatizar o posicionamento dos estores, contribuindo assim para uma maior eficiência no consumo energético (iluminação e climatização).

2. Problema

Os estores da fachada do edifício (Figura 1) têm já alguma idade, e por isso para além de permitirem a posição vertical deste, também é possível colocá-lo com uma determinada inclinação, não existindo posição intermédia entre estas.

![Figura 1 – Fachada do edifício](image1.png)

Como tal achou-se interessante fazer um estudo relacionando, a abertura do estore e a sua inclinação, visto que é possível incliná-lo, mas este estar totalmente fechado.

Apenas foi possível fazer esse estudo para os estores do piso intermédio, pois não se teve acesso às janelas do piso superior, de forma a colocar os estores nas posições necessárias.

3. Metodologia

A abordagem escolhida para conseguir definir se o estore está aberto ou fechado, foi a medição da área visível em cada uma das janelas, isto é, não coberta pelo estore. Assim, consoante a área visível, é possível determinar se cada estore se encontra fechado ou aberto. Para calcular os valores destas área foi necessário obter uma imagem com cada um dos estores totalmente fechado na vertical, que serviu como referência (Figura 2), assim como uma imagem da janela totalmente descoberta, de forma a saber qual a quantidade de pixéis existentes em cada uma delas.

![Figura 2 – Imagem da fachada que serve como referência](image2.png)

Foi necessário recorrer a diferentes ferramentas de manipulação de imagem para conseguir encontrar a quantidade de pixéis da área descoberta da janela. É possível ver no diagrama da Figura 3 o algoritmo elaborado para conseguir classificar a posição dos estores.

À imagem resultante da subtração da imagem original e da referência foi aplicada a binarização [1]. Para tal foi necessário encontrar o melhor valor de threshold, tendo sido feito de forma manual. No software Labview, é necessário inserir os valores de mínimo e máximo valor do pixel durante o threshold, tendo sido definido 25 e 150, respectivamente.

 Após a binarização, através da aplicação do filtro morfológico Open com um elemento estrutural em forma de quadrado com 7x7 pixeis [2], foi possível eliminar os pixeis que não traziam informação relevante.

Visto que a incidência solar varia com o passar das horas, e tendo sido a imagem de referência retirada com uma determinada incidência, é necessário definir uma região de interesse da subtração das imagens (figura 4), para não sermos influenciados, por exemplo, pelos pixeis da parede. A imagem com a região de interesse já definida é aplicada então uma máscara para eliminar todos esses pixeis sem interesse para a classificação do estore, como referido anteriormente.

![Figura 4 – Definição da região de interesse em cada uma das janelas](image3.png)

Após estes processos serem aplicados, foi analisada a dimensão da área de cada janela, através do número de pixeis correspondente e a inclinação do estore através do método de Hough [3] (Figura 5).

![Figura 5 – Deteção da inclinação dos estores pelo método de Hough](image4.png)
Quanto à classificação da posição dos estores, foi feita a distinção entre “estore direito” e “estore inclinado”, e quanto à sua abertura, foram criadas as classes “aberto”, “meio-aberto” e “fechado”. Para fazer tal classificação, foi realizada um processo manual de atribuição de intervalos de números de pixéis para cada uma das janelas, visto a imagem não ser frontal, e todas as janelas acabarem por ser “diferentes” umas das outras. Com o intervalo definido para cada uma delas, foi possível classificar de acordo com as posições anteriormente referidas.

Desta forma foi criado um mostrador referente a cada janela (Figura 6), de forma a se saber qual a classificação referente a cada uma delas.

### 4. Resultados

A maior dificuldade sentida na classificação dos estores refere-se ao facto de a luminosidade variar ao longo do dia, e de dia para dia. Nas figuras seguintes é possível avaliar os resultados obtidos para duas imagens retiradas em diferentes exposições solares.

![Figura 6 – Mostrador da classificação dos estores](image)

**Figura 7 – Exemplo do resultado obtido para uma incidência solar.**

**Figura 8 - Exemplo do resultado obtido para uma incidência solar distinta.**

O algoritmo de classificação da posição dos estores é bastante rápido, o que o torna executável no caso de haver uma grande quantidade de imagens para processar.

### 5. Conclusões

Através deste trabalho foi possível concluir que as potencialidades da visão computacional são muito vastas e que é possível implementar soluções a nível de classificação e decisão, baseadas nos resultados obtidos através de imagens. Desta forma, e tendo em conta o banco de imagens usadas para testar o programa desenvolvido, conclui-se que a classificação deste é eficaz.

Conseguiu-se desenvolver um software que independentemente da incidência solar nos estores, é capaz de classificar a sua posição e inclinação.

Este trabalho foi realizado com o objetivo dos seus resultados serem utilizados para um projeto de automatização de estores baseado nas condições climatéricas. Daí concluir-se que este programa apresenta resultados fundamentais no desenvolvimento deste projeto.

### 6. Trabalho futuro

Neste trabalho foram definidos três intervalos de classificação da posição do estore. No entanto para se obter resultados mais precisos que resultariam numa maior eficácia na implementação de domótica, seria fundamental definir um maior número de intervalos de classificação, de forma a se conseguir identificar um posicionamento mais preciso do estore. Assim em vez de apenas se ter as posições “aberto”, “meio-aberto” e “fechado”, poderia ter-se a percentagem aberta do estore, e conseguir definir assim um maior número de posições deste.

### Agradecimentos

Começo por agradecer ao LNEC o interesse pelo desenvolvimento deste trabalho, assim como a sua sugestão, facultando-me o acesso às fotografias retiradas à fachada do edifício.

Este trabalho foi suportado pela FCT, através do IDMEC, LAETA

### 7. Referências


Abstract

This paper addresses the image mosaicking problem using images from an UAV. First, the Harris-Laplace method is used to find scale-invariant keypoints, after which, using the same method as in SIFT (Scale-Invariant Feature Transform), it is possible to compute a descriptor for each keypoint. Ransac, together with the DLT (Direct Linear Transform) algorithm, is subsequently used to robustly estimate the homography between two images.

Using virtual images from a simulator of an UAV flight, solutions are presented to increase the quality of mosaics by exploring utilization of both image and metadata based methods in the mosaicking process. Results show the importance of decreasing the amount homographies multiplications and implementation of solutions to diminish numerical error propagation.

1 Introduction

Aerial photography has always been considered as an excellent form to gather information about the Earth’s surface. At the beginning, images were simply manually pieced together, but after the first satellites started sending pictures back to Earth, the need for automatic stitching operations between images increased.

A lot of research has been done in image mosaicking methods over the last few years. Using a low cost UAV (Unmanned Aerial Vehicle), the work presented in this paper falls within the scope of activities of LAETA (Laboratório de Energia, Transportes e Aeronáutica), namely in the Firecamp project which tries to assess fire related safety in campgrounds and trailer parks in Portugal.

Several difficulties and challenges are reported in literature [1][3] when images are obtained from low-flying, small-scale UAVs, particularly when using homographies. Common issues are: non-planar surfaces, low altitude, non-nadir views, inaccurate data from UAV coupled sensors and lens distortions.

The goal of this paper is to evaluate the impact of errors in the mosaicking process and present solutions to create geometrically and visually accurate mosaics. For this purpose, mosaics of the developed methods and an analysis of the error committed by different approaches are presented.

In the next section the followed methodology to build mosaics is presented. Section 3 describes the developed solutions to improve mosaics’ quality. Sections 4 and 5 are dedicated to the discussion of results and conclusions.

2 Mosaicking process

In order to create a mosaic it is necessary to find the relation among each of the individual images from which it is built, a procedure known as image registration.

According to Zisserman et al. [11], image registration consists on matching a set of images that portray a similar part of the ground, i.e., have an overlapping region. This can be made by using different methods, listed in [5] and [6], being of special interest for this paper the automatic ones: image and metadata based. The first method depends of accurate detection and matching of keypoints in different images, while the second is influenced by sensor inaccuracies; thus, both are subject to errors.

In [2], [3], [4] and [7], at least one of these methods are used to match images to overcome the challenges of using UAVs to acquire images of the ground.

Largely implemented, the image based method, when opposed to the metadata based approach, produces seamless mosaics when there are variations of the UAV’s attitude. This is not guaranteed when metadata is used due to sensor inaccuracies. However, its implementation is more permeable to error propagation, which introduces distortion into a mosaic. In this paper, correction of the UAV attitude is always made using image based methods.

The algorithm implemented by [2] uses the Harris-Laplace corner detector [8] to find interest points in each image and the SIFT method to then assign a canonical orientation to these keypoints, and ultimately, describe them.

After being described, feature points on both images are matched, where, according to D. Lowe [9], the best applicant to a match is the one with the minimal Euclidian distance of the invariant descriptor. The resulting set of points, or putative correspondences, are just an initial estimate fallible to mismatches. To contour this problem, Ransac algorithm [10] is used to eliminate mismatches, which combined with DLT algorithms, enables to estimate a homography between two images.

Admitting a sequence of images where each one overlaps the previous and the following, and assuming that only image based methods are used, to relate each image to the reference (e.g. the first) it is necessary to calculate composed homographies – the result of multiplying more than one estimated homographies. These homographies are the relation between a given image k and the reference.

Using the simulator developed in [2], all mosaics shown in this paper, except that of Fig.6, are built from virtually acquired images whose overlapping percentage is always 50% [1][2] – in both horizontal directions – and have a perfect nadir-view of the ground. This allows simulating the best possible conditions to the mosaicking process and, establishes a comparison between methodologies as the same acquired images are used to create mosaics using different methods.

3 Mosaicking methodologies

Following four different solutions for the mosaicking problem gives different results when it comes to mosaic quality, especially when the number of matrix multiplications is reduced.

Building a mosaic of 60 images using solely image methods, while following the sequence of photographs as taken during the flight is shown in Fig.1. The observed distortion is a consequence of multiplying 59 homographies to place every image in the reference frame of the first.

A mosaic of real images, acquired with an UAV overflowing the Coja campground, built using this method is shown in Fig.7, where the effects of distortion are also noticeable.

Fig.1: Mosaic of 60 images using original sequence.

Shown in Fig.3, is the result of using the same images as in the previous case, but in a reorganized order to reduce the maximum number of homography multiplications to 17. Assuming images organizes in a matrix of 4 lines and 15 columns as shown in Fig.2, images in the first column are placed in the mosaic and only after that, those in lines are stitched to the first image of each line: sequentially, images 1 and 30 are registered, then images 31 and 30 and finally 60 and 31. After this, images within lines are stitched to the first image of each line: image 2 to 1, 29 to 30, 32 to 31 and 59 to 60, and so on until each line’s last image.

Fig.2: Matricular organization of images in columns and lines

Following the same method which gives the mosaic of Fig.1, it is presented a hybrid method where both image and metadata are used.
Camera’s translation information from metadata is used in the homography matrix correspondent parameters, while the rotational ones are the result of image based registration. The outcome is shown in Fig.4.

A different solution that neither use image nor metadata based information is the placement of Ground Control Points (GCP) near the surface. These are simply GPS markers with high accuracy, whose relative position from each other is exactly known, as schematically shown in Fig.5. Placing a minimum of markers is related with the least amount of points from which a planar homography can be computed: one point for each two degrees of freedom.

Using these markers, after an interval where only image based methods are used to stitch photographs, it is possible to stop error propagation by using a homography calculated with the exact information from the GCP. Fig.6 shows a mosaic built using this method, where markers are in intervals of two images, nearly 560 meters apart.

4 Error analysis

It is possible to observe the evolution of error propagation due to homography multiplication. Computing the deviation, in pixels, of the estimated location of a matched pair on the second image using estimated and composed homographies, when the true location of that point and its match on the first image are known, allows comparing the presented methods. The result is shown in Fig.8.

Numerical errors are assumed to be the source of the evolution of Fig.8, however, its results show that with the increase of homography multiplications, the error committed increases, as expected after visual inspection of Fig.1.

5 Conclusions and future work

Results presented in this paper show that even for virtually perfect conditions, image based methods tend to compromise mosaic quality when the number of homography multiplications needed to create a mosaic is at least more than 17. Results also show that implementing a hybrid method for virtual images offers visually satisfactory results, as well as using GCP to compute exact homographies between non-consecutive images.

As future work, building mosaics of images applying the presented methods, using real images acquired with a GoPro camera and a regular webcam will be addressed. Results of creating mosaics from GoPro photographs, where lens distortion is corrected or only the central part of such images is considered in the mosaicking process will also be compared.

Acknowledgments

This work was partially funded by the Project PEst-OE/EME/LA0022/2011, LAETA Unit (IDMEC-Pole IST) of the Foundation for Science and Technology (Fundação para a Ciência e Tecnologia).

References

SignalBIT Framework: Principles and Applications

Ana Priscila Alves
AnaPriscila.Alves@lx.it.pt
Hugo Silva
hsilva@lx.it.pt
André Lourenço
alourenco@deetc.isel.ipl.pt
Ana Fred
alfred@lx.it.pt

1 Instituto de Telecomunicações
Instituto Superior Técnico
Avenida Rovisco Pais, 1, 1049-001 Lisboa, Portugal
2 Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal

Abstract

Biosignals are being used in a large variety of applications, in areas such as sports, healthcare, quality of life, among others. A major problem, in the context of biosignal research, is the rapid prototyping of end-user applications. Typically, there are either highly flexible scientific computing tools (characterized by long learning curves and limited user interface design capabilities), or custom developed and application-specific tools (characterized by poor cross-platform compatibility). Either case has several shortcomings, which generally result in high overheads in terms of development time whenever new features are needed. In this paper we describe the main principles and a set of applications of SignalBIT, a versatile, flexible, and extensible software framework for rapid prototyping of end-user applications, specifically targeted at biosignal acquisition and post-processing. The SignalBIT framework build on the advantages of combining web technologies with the Python programming language.

1 Introduction

The widespread use of biosignals in a large variety of applications demonstrates that it is an increasingly growing field of interest both in academia and industry [2], however in the context of biosignal research, a major bottleneck is the rapid prototyping of high impact end-user applications. In the overall, there is a need for software architectures and tools that can overcome the limitations of existing approaches.

Web technologies have changed many established paradigms in software development, by allowing the creation of user-friendly applications with multi-tier architecture, without requiring complex development or deployment procedures. Since they only require a compatible web browser, there is a high cross-platform capability; and with the advent of HTML5, programmers can create rich interactive environments natively within browsers, allowing the development of tools that can more easily meet users needs.

In this work we describe the principles behind SignalBIT, an innovative software framework proposed by our group [1], designed for rapid application development, and that can be used in a large variety of scenarios within biosignal research. The remainder of the paper is organized as follows: Section 2 describes the SignalBIT framework; Section 3 provides a few examples of applications of the framework; and finally, Section 4 highlights the main conclusions and future work perspectives.

2 Framework

The software framework was designed under a Model-View-Controller (MVC) approach, building on the original work by [4], where the presentation, control and model layers are decoupled and interchangeable, allowing a high degree of versatility. Figure 1 describes the architecture here proposed, where we have a backend implemented in the Python programming language, accessible over Websockets through a transparent and asynchronous Message Passing Protocol (MPP), a front-end implemented by means of web languages, and a controller layer responsible for the message transmission between back and front-end.

The backend uses the Simplest Websocket Server in the West (sWSw)\(^1\) with a custom callback function. Once the backend is launched it listens for inbound connections; when a client successfully establishes a connection, the backend waits for messages to arrive through the input stream and executes the callback function for each message. Our callback function assumes that the messages arriving from the input stream correspond directly to Python commands. As such, when a command is received through the input stream, it is evaluated in the Python backend and the result of its execution is stored.

Once the command has completed, the backend sends a message to the client through the output stream, formatted as a command with the same name as the one that was received, but in which the arguments are replaced the result of the backend execution. The arguments both on the inbound and outbound commands are formatted using JSON, a standard data-interchange format based on dictionary-type structures, which is supported in most modern programming languages and also the native data representation format in Javascript.

The front-end was designed considering that it must provide intuitive and interactive interfaces. Web browsers currently offer the versatility of being used in all operative systems, combined with the possibility to create a rich user interface experience with relative ease of layout design and formatting; HTML is the base technology, which is responsible for modelling the web page structure. Alongside with this technology there are the Cascading Style Sheets (CSS), which controls the style and layout of the web page. JavaScript provides a comprehensive set of functions for user-interface event handling, interaction logic, and browser-side computing.

![Figure 1: SignalBIT framework architecture.](image)

![Figure 2: Code snippets of the Python backend and a web-based frontend.](image)
The link between the Python backend and the web-based front-end, is the controller layer implemented in JavaScript, which detects any event occurring in the user interface, and translates that to executable functions in the backend. The HTML, CSS and Javascript approach also has the advantage of making the presentation layer highly reusable as most of the commonly used programming languages have UI components to load and render HTML/CSS (e.g. the WebView component on the Java SDK for Android).

Figure 2 shows simplified code snippets of the Python backend and a web-based front-end. In this example the client sends the command `upper('hello world')` to the Python backend when the Websocket connection is established, the backend executes the command and responds to the client with the command `upper('HELLO WORLD')` (the uppercase conversion of the original string).

3 Applications

The SignalBIT framework lays on the independence between the model and view layers, which gives a high versatility in the construction of applications. We will describe here two examples designed under the architecture proposed: a) SignalBIT, a platform for real-time biosignal recording and visualization, and b) FlowerBIT, a platform for monitoring a flower.

3.1 SignalBIT

The SignalBIT [1] platform developed is a set of software components for data acquisition, visualization and recording, which includes a framework supported by the Python programming language for online and offline data handling, and an user-friendly graphical user interface (GUI). Figure 3 shows the overall interface that we entitled SignalBIT. A global overview plot on the bottom of the work area provides a 1 minute summary of all the analog channels being acquired. The graphics on the center of the screen show the individual channels and the separator between the two areas indicates the digital inputs state and allows the user to change the digital outputs.

![Figure 3: SignalBIT application user interface](Image)

The backend in Python is responsible for the connection with a BITalino device [3] via Bluetooth, and for all the data processing and storage. The main processes executed in this layer are: a) the access to the file system, or retrieve previously stored session configuration and settings; b) acquisition using the Bluetooth socket (which in our case is supported by the PyBluez API), and receiving the data from the BITalino in real-time; and finally c) format the data as a JSON object interpretable by the PlotJS API. In the front-end layer, the user can: a) start/stop acquisitions; b) choose the acquisition configurations, such as sampling rate, number of channels to acquire and the device to communicate to; and c) save the recorded data for later processing.

When an acquisition starts, the data being acquired is displayed in real-time and the user is presented with a set of options to change the zoom in the amplitude and time scale, and also change the offset. Since this architecture decouples the logic layer from the visualization, any researcher can recycle the back or front-end layer and change it to fit its applications needs.

3.2 FlowerBIT

The simplicity implied by this framework can be extended to any kind of application, even if it is not related to biosignals. For instance, FlowerBIT is a platform for monitoring a flower, which collects information as soil moisture, temperature, light and ambient humidity, to check the plant “health” state. In this application, each sensor is connected to a BITalino microcontroller and the data is sent to the base station via USB. The backend in Python is responsible for collecting the data and send each value to the web front-end. Figure 4 shows the front-end, where each parameter being measured is displayed.

![Figure 4: FlowerBIT application user interface](Image)

This example shows the potential use of this framework in data visualization and monitoring. However, it can be applied in many scenarios, allowing a non-proficient programmer to rapidly develop an application.

4 Conclusion and Future Work

In this paper we described the principles and applications of the SignalBIT framework for rapid application development, which is not only suitable to acquire biosignals in real-time, but also in multiple dimensions of academic programs and workshops, opening new research horizons and prospects. Our work was geared towards the creation of a simplified albeit flexible and scalable framework that can enable rapid application development of high impact software prototypes involving real-time biosignal acquisition and analysis. We believe that this architecture has a good compromise solution between usability and performance tradeoffs. Future work will focus on further optimizing the performance of the core operations performed by the Model, and on benchmarking and evolving our framework to a web-based format.

5 Acknowledgment

This work was partially funded by Fundação para a Ciência e Tecnologia (FCT) under the grants PTDC/EEI-SII/2312/2012, SFRH/BD/65248/2009 and SFRH/PROTEC/49512/2009.

References


Correction of Geometrical Distortions in Bands of Chromatography Images

Bruno Moreira\textsuperscript{1,2} 
António Sousa\textsuperscript{1,3} 
Ana M. Mendonça\textsuperscript{1,2} 
Aurélio Campilho\textsuperscript{1,2} 

\textsuperscript{1} Instituto de Engenharia Biomédica, Universidade do Porto 
\textsuperscript{2} Faculdade de Engenharia, Universidade do Porto 
\textsuperscript{3} Instituto Superior de Engenharia, Instituto Politécnico do Porto

Abstract
This paper presents a methodology for correcting band distortions in Thin-Layer Chromatography (TLC) images. After the segmentation of image lanes, the intensity profile of each lane column is spatially aligned with a reference profile using a modified version of the Correlation Optimized Warping (COW) algorithm. The proposed band correction methodology was assessed using 105 profiles of TLC lanes. A set of features for band characterization was extracted from each lane profile, before and after band distortion correction, and was used as input for three distinct one-class classifiers aiming at band identification. In all cases, the best results of band classification were obtained for the set lanes after band distortion correction.

1 Introduction
Thin-Layer Chromatography (TLC) lanes contain bands that are often affected by geometric distortions resulting from the sample development process. Band deformation can modify significantly the morphology of the lane intensity profile and bias the extraction of features for band characterization, such as band area and band peaks. In extreme circumstances, the distortion can even prevent the occurrence of a band in the intensity profile either by its integration into another band or into the lane background. An example of a typical TLC image can be observed in Fig. 1, where several bands with and without distortion are marked by the continuous and dashed line rectangles, respectively.

![Figure 1: Region of interest of a TLC image. The continuous and dashed line rectangles mark bands with and without distortion, respectively.](image)

Methodologies to deal with band distortion in images similar to TLC images, such as electrophoresis, have been proposed before. The band straightening algorithm developed for the GASepo system \cite{1} significantly reduces band geometric distortions in Epo images. The main idea behind this methodology is an automatic search for an optimum vertical shift in all image columns that leads to a maximum reduction of local geometric distortions. In \cite{2} an algorithm is proposed for removing band distortion in electrophoretic gels using a gradient filter to estimate band orientations. Bands are smoothed and interpolated by robust regression, and integrated to find the unwarping transform. In \cite{3} a two-step method is proposed. Firstly, a simple physicochemical model is formulated and applied for warping each gel image to correct for leakage across the sides during electrophoresis. Secondly, the alignment is obtained by maximizing a penalized likelihood criterion. The likelihood measures the similarity between aligned images and the penalty is a distortion criterion.

This paper describes a methodology for improving the quality of digital images of TLC plates consisting in the correction of band distortions, thus facilitating the subsequent characterization of bands associated with relevant biomarkers. In our application, we are interested in the identification of Fabry disease (FD). For this purpose, we will focus on the detection a specific biomarker, the globotriaosylceramide (Gb3).

2 Methodology
The proposed method for band distortion correction is a modified version of the Correlation Optimized Warping (COW) algorithm \cite{4} that overcomes some of the problems that came out from the application of the original solution to TLC chromatograms of human urine samples.

The COW algorithm was introduced by Nielsen et. al. \cite{4} aiming at the correction of misalignments and shifts between chromatographic 2-D profiles by aligning a sample profile with a reference one by enlarging or compressing sample segments using linear interpolation. An equal number of segments of length $f$ is specified in the reference and sample profiles. The maximum shift in the length of a sample segment is controlled by the slack parameter $\tau$.

The modified COW algorithm presented by Skov et al. in \cite{5} includes an automatic method for the selection of the reference chromatogram and an optimization of the $f$ and $\tau$ parameters through the maximization of the warping effect. This quantitative measure combines two functions with equal weight, the simplicity value and the peak factor $\delta$. The simplicity value allows to measure how well aligned a set of profiles is. The peak factor tries to guarantee that no strong compression (or enlarging) is required to align the profiles.

The herein proposed methodology for correcting deformations in TLC bands is based on the improved COW algorithm developed by Skov et al. \cite{5}. Our algorithm will be applied in TLC images with a different objective as, instead of aligning different chromatographic profiles, we want to correct band distortions such as smiles, frowns and slopes. These three types of distortion are illustrated in Fig. 2.

![Figure 2: Typical distortions found in TLC bands. (a) Smiley. (b) Frown. (c) Slope.](image)

Since the original algorithm in \cite{5} could not achieve satisfactory results, it was adapted to cope with the limitations that were found. The main problems related to the original algorithm are depicted in Fig. 3. Fig. 3a presents the original images and Fig. 3b shows the images after band correction with the approach described in \cite{5} (using the original code provided by the authors): in the top row the contour of an original flawless band becomes irregular; in the middle row, band positions are modified by the automatically selected reference and the column alignment is spread to other lane regions; in the bottom row, the segment boundary separates the band and the deformation correction is unsuccessful. Fig. 3c presents the results obtained with the application of the herein proposed methodology, which is not affected by the described problems.

![Figure 3: Correction of band distortions. (a) Original images. (b) Results obtained with the original algorithm described in [5]. (c) Results obtained with our approach.](image)

The lines in the TLC plate image are segmented using the methodology described in \cite{6}. Each lane column is represented as an intensity profile to be aligned with the lane reference. The product of the correlation...
coefficients between all individual profiles was suggested in [5] as an appropriate solution for selecting the reference column from the lane. However, this approach is not included in our algorithm as in a deformed band the central columns are normally those that better describe the correct position. So, the reference profile, $R(y)$, for a lane image $I(x,y)$ is obtained by

$$R(y) = \frac{1}{2S+1} \sum_{i=-S}^{S} I \left( \frac{N}{2} + i, y \right), \quad y = 1 \cdots M$$ (1)

where $M$ is the number of rows and $N$ is the number of columns in the lane image. $S$ is the number of columns on each side of the lane centre that is included in the averaging of the reference profile. In this work, a value of $S=3$ was selected.

The profile partition into equal length segments as performed in COW can result in the separation of a single band in two segments to be disjointedly aligned, which makes the result highly dependent on the localization of segment limits. To overcome this problem, the implemented algorithm only applies the warping function in selected lane regions corresponding to groups of bands, identified as local maxima regions in the lane intensity profile. As a result of this region selection process, nearby bands are grouped together and aligned as a whole, but at the same time each group of bands is warped independently from all the others.

In our implementation, for calculating the warping effect value different weights are assigned to the simplicity and peak factor, being the simplicity weight three times higher than the peak factor weight. In band correction, a strong compression (or stretching) is often necessary due to the high displacement between the profiles, which inevitably creates a low peak factor. This overly influences the warping effect and prevents the alignment of the columns further away from the centre. Thus, it was decided that the simplicity value should have a greater impact in the choice of the correct values for $l$ and $t$.

3 Results

The application of the proposed methodology is exemplified in Fig. 4. A sample lane from Fig. 1 (lane 6) is shown in Fig. 4a. The result of the aligning process on the lane bands can be observed in Fig. 4b, where the “smiley” effects have disappeared. Figs. 4c and 4d show the reference and a column intensity profile, before and after band correction, respectively. The methodology allows profile displacement and peak alignment, as can be observed in Fig. 4d. The mean lane profile that is obtained after the column projection over the lane development direction is presented in Figs. 4e and 4f, before and after the band correction process, respectively. The middle band is visible in the corrected lane profile and the peaks present a higher intensity value than in the original profile.

![Figure 4: Illustration of the results obtained with the band correction methodology. (a) Original lane of a TLC image. (b) Lane reconstruction after band correction. (c-d) Reference profile (continuous line) and profile obtained from a column (dashed line) before alignment (left) and after alignment (right). (e-f) Integration of lane information before band correction (left) and after band correction (right).](image)

The quantitatively evaluation of band correction was achieved by identifying the Gb3 band in specific regions of the profile, before and after the application of the methodology. A set of 105 regions (53 bands and 52 empty regions) was used. Band identification was based on a set of four features extracted from each lane profile, before and after band distortion correction, and was used as input for three distinct one-class classifiers [8]. The mean and standard deviation values, obtained after 100 runs of 10-fold cross validation method, are shown in Table 1. Both the mean error and variance decreased after band correction using the proposed algorithm. The random behaviour of the original COW prevented the achievement of satisfactory results.

<table>
<thead>
<tr>
<th>Table 1. Mean error and standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
</tr>
<tr>
<td>Before band correction</td>
</tr>
<tr>
<td>Original COW [5]</td>
</tr>
<tr>
<td>Proposed algorithm</td>
</tr>
</tbody>
</table>

4 Conclusion and Future Work

A new methodology for the automatic correction of band deformations in chromatography images based on the Correlation Optimized Warping (COW) has been proposed. The method overcomes some limitations of the original algorithm when applied to TLC images. The advantages of band correction were validated both visually by inspecting the reconstructed lanes and quantitatively by using the same set of features, measured before and after the application of the methodology, to identify Gb3 bands. The improvement of characteristics related to the presence of bands, such as the peak intensity values, lead to an increase in the number of correct identifications.

Acknowledgements

The authors would like to thank the Portuguese Foundation for Science and Technology (FCT) (Ref. PCT PTDC (SAU/BEB/100875/2008) and FEDER (Ref. FCOMP-01-0124-FEDER-010913) for co-funding this research project.

References


RecPad 2013 Page 56 of 147
A novel sparsity and clustering regularization

Xiangrong Zeng
Xiangrong.Zeng@lx.it.pt
Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt

Abstract

We propose a novel SPARSity and Clustering (SPARC) regularizer, which is a modified version of the previous octagonal shrinkage and clustering algorithm for regression (OSCAR), where, the proposed regularizer consists of a $K$-sparse constraint and a pair-wise $\ell_\infty$ norm restricted on the $K$ largest components in magnitude. The proposed regularizer is able to separably enforce $K$-sparsity and encourage the non-zeros to be equal in magnitude. Moreover, it can accurately group the features without efficiently computed based on that of the latter, allowing using proximal splitting algorithms to solve problems with SPARC regularization. Experiments on synthetic data and with benchmark breast cancer data show that SPARC is a competitive group-sparsity inducing regularizer for regression and classification.

1 Introduction

In recent years, much attention has been paid not only to sparsity but also to structured/group sparsity. Several group-sparsity-inducing regularizers have been proposed, including group LASSO (gLASSO) [9], fused LASSO (fLASSO) [7], elastic net (EN) [11], octagonal shrinkage and clustering algorithm for regression (OSCAR) [5], and several others, not listed here due to space limitations (see review in [1]). However, gLASSO (and its many variants and descendants [1]) require prior knowledge about the structure of the groups, which is a strong requirement in many applications, while LASSO depends on a given order of variables; these two classes of approaches are thus better suited to signal processing applications than to variable selection and grouping in machine learning problems, such as regression or classification (where the order of the variables is often meaningless). In contrast, EN and OSCAR were proposed for regression problems and do not rely on any of the ordering or variables or knowledge about group structure. The OSCAR regularizer (shown in [10] to outperform EN in feature grouping) is defined as

$$
\theta_{\text{OSCAR}}(x) = \lambda_1 \|x\|_1 + \lambda_2 \sum_{i,j} \max \{ |x_i|, |x_j| \}
$$

where $\lambda_1$ and $\lambda_2$ are non-negative parameters (which, in practice, can be obtained, for example, by cross validation) [10]. The $\ell_1$ norm and the pairwise $\ell_\infty$ penalty simultaneously encourage the components to be sparse and equal in magnitude, respectively. However, it may happen that components with small magnitude that should be shrunk to zero by the $\ell_1$ norm are also penalized by the pairwise $\ell_\infty$ term, which may prevent accurate grouping; moreover, components with large magnitude that should simply be grouped by the pairwise $\ell_\infty$ norm are also shrunk by the $\ell_1$ norm (see Figure 1). In this paper, to overcome these drawbacks, we propose the SPARSity-and-Clustering (SPARC) regularizer, where the cardinality of the support of the solution is restricted and the pairwise $\ell_\infty$ penalty is applied only to the non-zero elements (see Figure 1). We also show how to compute the proximity operator of the SPARC regularizer (although it is non-convex), which allows using proximal splitting algorithms to problems with this regularizer.

2 Proposed Formulation and Approach

A linear regression problem (with design matrix $A \in \mathbb{R}^{n \times p}$) under SPARC regularization is formulated as

$$
\min_{x} \frac{1}{2} \|y - Ax\|^2 + \theta_{\text{SPARC}}(x) = \min_{x} \frac{1}{2} \|y - Ax\|^2 + \lambda_1 \|x\|_1 + \lambda_2 \sum_{i,j} \max \{ |x_i|, |x_j| \}
$$

where $\lambda_1$ and $\lambda_2$ are non-negative parameters (which, in practice, can be obtained, for example, by cross validation) [10]. The $\ell_1$ norm and the pairwise $\ell_\infty$ penalty simultaneously encourage the components to be sparse and equal in magnitude, respectively. However, it may happen that components with small magnitude that should be shrunk to zero by the $\ell_1$ norm are also penalized by the pairwise $\ell_\infty$ term, which may prevent accurate grouping; moreover, components with large magnitude that should simply be grouped by the pairwise $\ell_\infty$ norm are also shrunk by the $\ell_1$ norm (see Figure 1). In this paper, to overcome these drawbacks, we propose the SPARSity-and-Clustering (SPARC) regularizer, where the cardinality of the support of the solution is restricted and the pairwise $\ell_\infty$ penalty is applied only to the non-zero elements (see Figure 1). We also show how to compute the proximity operator of the SPARC regularizer (although it is non-convex), which allows using proximal splitting algorithms to problems with this regularizer.

![Figure 1: Demonstration of different regularizers](image)

The key observation that allows computing $\text{prox}_{\lambda K \text{SPARC}}(x)$ is

$$
x \in \Sigma_K \Rightarrow \lambda K \text{SPARC}(x) = 0
$$

where $x_k \in \mathbb{R}^5$ is the sub-vector of $x$ indexed by an index subset $S \subseteq \{1, ..., p\}$. Combining this with properties of proximity operators and ideas from [6] allows showing (naturally, details are omitted here) that $x = \text{prox}_{\lambda K \text{SPARC}}(x)$ can be computed as follows:

$$
\Omega_k(x) = \text{prox}_{\lambda K \text{SPARC}}(x_k), \quad z_{\Omega_k}(x) = 0
$$

where 0 is a vector of zeros, $\Omega_k(x) = \{1, ..., p\} \setminus \Omega_k(x)$, and $\text{prox}_{\lambda K \text{SPARC}}(x)$ can be obtained using the algorithm proposed in [10]. Therefore, we can solve (1) by proximal splitting algorithms, such as FISTA [3], TwIST [4], or SpaRSA [8], which is the algorithm adopted in our experiments. SpaRSA (which stands for sparse reconstruction by separable approximation [8]) is a fast proximal splitting algorithm, based on the step-length selection method of Barzilai and Borwein [2]. Its application to SPARC leads to the following algorithm:

**Algorithm SpaRSA for solving (1)**

1. Set $k = 1$, $\eta > 1$, $\alpha_0 = \alpha_{\text{min}} > 0$, $\alpha_{\text{max}}$, and $x_0$.
2. $x_0 = \mathbf{0}$ and $\eta_k = \frac{1}{\alpha_k} (A^T (A_k x_k + y) - \eta_k) / \sqrt{n}$.
3. $x_k = \text{prox}_{\lambda K \text{SPARC}}(x_0)$.
4. repeat
5. $\alpha_k = \|x_k - x_{k-1}\|_\infty$, $\alpha_k = \max \{ \alpha_k, \min \{ \alpha_k, \alpha_{\text{max}} \} \}$
6. until $x_{k+1}$ satisfies an acceptance criterion.
7. $k = k + 1$.
8. until some stopping criterion is satisfied.
3 Experiments

In this section, we report results of experiments with synthetic data and with the breast cancer benchmark data, aimed at comparing the SPARC with the LASSO, EN and OSCAR. In order to measure their performances, we employ the following six metrics defined on an estimate $e$ of an original vector $x^*$:

- Mean absolute error: $\text{MAE} = \|Ax^* - e\|_1$;
- Mean square error: $\text{MSE} = \|Ax^* - e\|_2^2$;
- Selection error rate: $\text{SER} = \|\|x^*\| - \|e\|\|_1/p$;
- Degrees of freedom (DoF): the number of unique non-zero coefficients of $e$;
- Classification accuracy (CLA): the number of correct classifications of $e$;
- Number of non-zero features (NNZ).

3.1 Synthetic data

we consider a regression problem where $y = Ax^* + w$, where the true parameters

$$x^* = \left[\begin{array}{c} 3 \\ \vdots \\ 3 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 3 \end{array}\right]^{T}$$

and the design matrix $A$ is generated as

$$a_i = z_1 + \epsilon^n, z_1 \sim N(0, 1), i = 1, \ldots, 5;$$
$$a_i = z_2 + \epsilon^n, z_2 \sim N(0, 1), i = 6, \ldots, 10;$$
$$a_i = z_3 + \epsilon^n, z_3 \sim N(0, 1), i = 11, \ldots, 15;$$
$$a_i \sim N(0, 1), i = 16, \ldots, 40$$

where $\epsilon^n$ are independent identically distributed $N(0, 0.16), i = 1, \ldots, 15$. And then $A = [a_1, a_2, \ldots, a_{40}]^{T}$ is further normalized, the noise variance of $w$ is 0.01. The number of samples for training, cross validation and testing are 20, 40 and 200, respectively. Notice that it is an ill-posed training problem, since the number of samples is less than the dimension of $x$ (20 < 40).

![Figure 2: Recovered parameters by different algorithms](image)

From Figure 2 and Table 1, the SPARC outperforms the LASSO, EN and OSCAR, showing it is a promising approach to feature selection and grouping in regression.

3.2 Breast cancer data

In this section, we report experiments with the benchmark breast cancer data, which contains 8141 genes in 295 tumors, where 300 genes that are most correlated with the responses. 50%, 30% and 20% of the data are then randomly chosen for training, cross validation, and testing, respectively. The results averaged over 50 repetitions are show in Table 2. We can observe that SPARC is a competitive group-sparsity-inducing regularizer for classification in terms of CLA, and it is able to select features with lower degrees of freedom than LASSO, EN, and OSCAR.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>LASSO</th>
<th>EN</th>
<th>OSCAR</th>
<th>SPARC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>27.0677</td>
<td>29.4458</td>
<td>66.2390</td>
<td>25.7473</td>
</tr>
<tr>
<td>MSE</td>
<td>7.6614</td>
<td>7.6939</td>
<td>36.8120</td>
<td>5.2904</td>
</tr>
<tr>
<td>DoF</td>
<td>15.64</td>
<td>25.28</td>
<td>4.56</td>
<td>4.02</td>
</tr>
<tr>
<td>SER</td>
<td>14.50%</td>
<td>25.75%</td>
<td>8.95%</td>
<td>5.50%</td>
</tr>
</tbody>
</table>

Table 2: Results of the metrics on synthetic data

4 Conclusions

We have proposed the SPARcity and Clustering (SPARC) regularizer for regression and classification. We have shown that the proposed SPARC is able to separably enforce $K$-sparsity and encourage the non-zeros to be equal in magnitude, thus accurately grouping the features without parameter shrinkage, outperforming the LASSO, the elastic net, and the octogonal shrinkage and clustering algorithm for regression (OSCAR). Future work will involve considering faster algorithms to solve problems with SPARC regularization.

References

Exploiting Two-Dimensional Group Sparsity in 1-Bit Compressive Sensing

Xiangrong Zeng
Xiangrong.Zeng@lx.it.pt
Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt
Instituto de Telecomunicações
Instituto Superior Técnico
Lisbon, Portugal

Abstract

We propose a new approach, two-dimensional fused binary compressive sensing (2DFBCS) to recover 2D sparse piece-wise signals from 1-bit measurements, exploiting 2D group sparsity for 1-bit compressive sensing recovery. The proposed method is a modified 2D version of the previous binary iterative hard thresholding (2DBIHT) algorithm, where the objective function includes a 2D one-sided \( \ell_1 \) (or \( \ell_2 \)) penalty function encouraging agreement with the observed data, an indicator function of \( K \)-sparsity, and a total variation (TV) or modified TV (MTV) constraint. The subgradient of the 2D one-sided \( \ell_1 \) (or \( \ell_2 \)) penalty and the projection onto the \( K \)-sparsity and TV or MTV constraint can be computed efficiently, allowing the application of algorithms of the forward-backward splitting (a.k.a. iterative shrinkage-thresholding) family. Experiments on the recovery of 2D sparse piece-wise smooth signals show that the proposed approach is able to take advantage of the piece-wise smoothness of the original signal, achieving more accurate recovery than 2DBIHT. More specifically, 2DFBCS with the MTV and the \( \ell_2 \) penalty performs best amongst the algorithms tested.

1 Introduction

We focus on recovering a sparse piece-wise smooth two-dimensional (2D) signal (an image) \( X \in \mathbb{R}^{N \times L} \) from 1-bit measurements,

\[
Y = \text{sign}(AX + W),
\]

where \( Y \in \{-1, 1\}^{M \times L} \) is the observed matrix of measurements, sign is the element-wise sign function, which returns \( +1 \) for positive arguments and \( -1 \) otherwise, \( A \in \mathbb{R}^{M \times N} \) is the known sensing matrix, and \( W \) is additive noise. Unlike in conventional compressive sensing (CS), 1-bit measurements have no information about the magnitude of the original signal \( X \). The goal is then to recover \( X \), but only up to an unknown and unrecoverable magnitude [2, 4].

Our innovations, with respect to 1-bit CS as proposed in [2], are twofold: a) we address the 2D case; b) more importantly, we introduce a new regularizer favoring both sparsity and piece-wise smoothness, which can be seen as a modified 2D version of fused lasso [6]. This new regularizer is the indicator of a union of convex subsets (total-variation balls) of the canonical subspaces, simultaneously enforcing sparsity and smoothness within each connected subset of non-zero elements. The rationale is that, when imposing smoothness and sparseness, smoothness should not interfere with sparsity, i.e., it should not be imposed across the transitions from zero to non-zero elements. The proposed regularizer promotes sparseness and smoothness and (although it is non-convex) has a computationally feasible projection, based on which we propose a modified version of the binary iterative hard thresholding (BIHT) algorithm [4].

2 2D Binary Iterative Hard Thresholding (2DBIHT)

To recover \( X \) from \( Y \), we first consider a 2D version of the criterion proposed by Jacques et al [4]

\[
\min_{X \in \mathbb{R}^{N \times L}} f(Y \odot AX) + \tau_\Sigma(X) \quad \text{subject to } \|X\|_2 = 1,
\]

where: the operation \( \circ \) denotes the element-wise (Hadamard) product; \( \tau_\Sigma(X) \) denotes the indicator function of set \( C \),

\[
\tau_\Sigma(X) = \begin{cases} 
0, & X \in C \\
\infty, & X \not\in C;
\end{cases}
\]

\( \Sigma_K = \{X \in \mathbb{R}^{N \times L} : \|X\|_0 \leq K\} \) (with \( \|V\|_0 \) denoting the number of non-zero elements in \( V \)) is the set of \( K \)-sparse \( N \times L \) matrices; \( f \) is one of the penalty functions defined next; and, \( \|X\|_2 = \sqrt{\sum_i X_i^2} \) is the Euclidean norm. To penalize linearly the violations of the sign consistency between the observations and the estimate [4], the barrier function is chosen as \( f(Z) = 2 \|Z\|_1 \), where \( Z = \min(\mathbb{Z}, 0) \) (with the minimum applied entry-wise and the factor 2 included for later convenience) and \( \|Y\|_1 = \sum_j |Y_{i,j}| \) is the \( l_1 \) norm of \( Y \). A quadratic barrier for sign violations (see [2]) is achieved by using \( f(Z) = \frac{1}{2} \|Z\|_2^2 \), where the factor \( 1/2 \) is also included for convenience. The iterative hard thresholding (IHT) [1] algorithm applied to (2) (ignoring the norm constraint during the iterations) leads to the 2DBIHT algorithm, which is a 2D version of the binary iterative hard thresholding (BIHT) [4]:

Algorithm 2DBIHT
1. set \( k = 0, \tau > 0, X_0 \) and \( K \)
2. repeat
   3. \( V_{k+1} = X_k - \tau \partial f(Y \odot (AX_k)) \)
   4. \( X_{k+1} = T_{\Sigma_K}(V_{k+1}) \)
   5. \( k \leftarrow k + 1 \)
   6. until some stopping criterion is satisfied.
   7. return \( X_k / \|X_k\|_2 \)

In this algorithm, \( \partial f \) denotes the subgradient of the objective (see [4], for details), which is given by

\[
\partial f(Y \odot (AX)) = \begin{cases} 
A^T \text{sign}(AX) - Y, & \ell_1 \text{ barrier}, \\
A^T (Y \odot (AX)), & \ell_2 \text{ barrier}.
\end{cases}
\]

Line 3 of this algorithm corresponds to a sub-gradient descent step (with step-size \( \tau \)), while line 4 performs the projection onto the non-convex set \( \Sigma_K \), which corresponds to computing the best \( K \)-term approximation of \( V \), that is, keeping \( K \) largest components in magnitude and setting the others to zero. Finally, the returned solution is projected onto the unit sphere to satisfy the constraint \( \|X\|_2 = 1 \) in (2). The versions of 2DBIHT with \( \ell_1 \) and \( \ell_2 \) objectives are referred to as 2DBIHT-\( \ell_1 \) and 2DBIHT-\( \ell_2 \), respectively.

3 2D Fused Binary Compressive Sensing (2DFBCS)

The proposed formulation essentially adds a new constraint of low (modified) total variation to the criterion of 2DBIHT (2), which encourages 4-neighbour elements to be similar, justifying the term “fused”.

3.1 2DFBCS with Total Variation

We first propose the following model:

\[
\min_{X \in \mathbb{R}^{N \times L}} f(Y \odot AX) + \tau_\Sigma(X) + \tau_T(X) \quad \text{subject to } \|X\|_2 = 1,
\]

where \( T_T = \{X \in \mathbb{R}^{N \times L} : \text{TV}(X) \leq \varepsilon\} \), with TV(x) denoting a discrete version of the total variation (TV) regularizer, given by

\[
\text{TV}(X) = \sum_{j=1}^{N-1} \sum_{i=1}^{L-1} \left( |X_{i,j+1,j} - X_{i,j,j}| + |X_{i,j,j+1} - X_{i,j,j}| \right),
\]

and \( \varepsilon \) is a positive parameter. In the same vein as 2DBIHT, the proposed algorithm is as follows:

Algorithm 2DFBCS-TV
1. set \( \tau > 0, \varepsilon > 0, K, \) and \( X_0 \)
2. repeat
   3. \( V_{k+1} = X_k - \tau \partial f(Y \odot (AX_k)) \)
   4. \( X_{k+1} = T_{\Sigma_K}(P_{\text{TV}(\varepsilon)}(V_{k+1})) \)
   5. \( k \leftarrow k + 1 \)
   6. until some stopping criterion is satisfied.
   7. return \( X_k / \|X_k\|_2 \)
5. \( k \leftarrow k + 1 \)
6. until some stopping criterion is satisfied.
7. return \( X / \| X \|_2 \)

In this algorithm, \( P_{T_k} \) computes the projection onto \( T_k \), which can be obtained by using the algorithm proposed by Fadili and Peyré [3]. The versions of the 2DFBCS-TV algorithm with \( \ell_1 \) and \( \ell_2 \) objectives are referred to as 2DFBCS-TV-\( \ell_1 \) and 2DFBCS-TV-\( \ell_2 \), respectively.

\[ \min_{X \in \mathbb{R}^{N \times L}} f(Y \otimes (AX)) + \lambda f_k(X) \]
subject to \( \| X \|_2 = 1 \),

where the set \( F_k \) requires a more careful explanation. As above, let \( \Sigma_k \) be the set of \( K \)-sparse N \times L matrices. Consider the undirected 4-nearest-neighbors graph on the sites of \( N \times L \) images, \( G = (\mathcal{N}, E) \), where \( \mathcal{N} = \{(i, j), i = 1, \ldots, N, j = 1, \ldots, L\} \) and \( \{(i, j), (i - 1) \} \in E \Leftrightarrow j \in X \). Given some image \( V \in \mathbb{R}^{N \times L} \), let \( \tilde{G}(V) = (\tilde{N}, \tilde{E}(V)) \) be the subgraph of \( G \) obtained by removing all the nodes corresponding to zero elements of \( V \) that is, \( (i, j) \in \tilde{N} \) if \( V(i, j) \neq 0 \), as well as the corresponding edges. Naturally, \( \tilde{G}(V) \) is not a connected graph; thus, define \( \{\tilde{G}(V), \ldots, \tilde{G}(V)\} \), as the set of the \( \tilde{G}(V) \) connected subgraphs of \( \tilde{G}(V) \), where \( \tilde{G}(V) = (\tilde{N}(V), \tilde{E}(V)) \). Define the normalized TV of the sub-image of \( V \) corresponding to each of these connected subgraphs as

\[ \text{TV}(\tilde{G}(V)) = |\tilde{E}(V)|^{-1} \sum_{(i, j), (k, l) \in \tilde{E}(V)} |V(i, j) - V(k, l)| \]

(assuming \( \| \tilde{E}(V) \| > 0 \) where \( V(i, j) \) is the subgraph indexed by \( \tilde{G}(V) \). Finally, the set \( F_k \) is defined as

\[ F_k = \{ X \in \Sigma_k : \text{TV}(\tilde{G}(X)) \leq \epsilon, \ k = 1, \ldots, \kappa(K) \}. \]

In short, \( F_k \) is the set of K-sparse images such that the normalized TV of each of its connected blocks of non-zeros doesn’t exceed \( \epsilon \). Notice that this is different from the intersection of a TV ball with \( \Sigma_k \), as considered in [5].

In the same vein as the 2DBIHT, we propose the following BIHT-type algorithm to solve (5):

**Algorithm 2DFBCS-MTV**

1. Set \( \tau > 0, \epsilon > 0, K \), and \( X_0 \)
2. repeat
3. \( V_{k-1} = X_{k-1} - \tau \partial f(Y \otimes (AX_k)) \)
4. \( X_{k+1} = P_{F_k}(V_{k+1}) \)
5. \( k \leftarrow k + 1 \)
6. until some stopping criterion is satisfied.
7. return \( X / \| X \|_2 \)

In this algorithm, line 3 is also a sub-gradient descent step, where \( \partial f \) is defined as (4) while line 4 performs the projection onto \( F_k \). Although \( F_k \) is non-convex, here we can briefly show that \( P_{F_k} \) can be computed as the follows: first, project onto \( \Sigma_k \), i.e., \( U = P_{\Sigma_k}(V) \); then, \( X = P_{F_k}(V) \) is obtained by projecting every connected group of non-zeros in \( U \) onto the \( \epsilon \)-radius normalized TV ball \( B_k \).

\[ B_k = \{ X \in \Sigma_k : \text{TV}(\tilde{G}(X)) \leq \epsilon \}, \]

for \( k = 1, \ldots, \kappa(K) \), i.e., \( X_{\Sigma_k} = U_{\Sigma_k}, \) for \( k = 1, \ldots, \kappa(K) \), and keeping the zeros of \( U \), i.e., \( X_{\tilde{G}(U)} = U_{\tilde{G}(U)} \). Finally, as in [4], projection onto the unit sphere (line 6) enforces the constraint in (5). The versions of the 2DFBCS-MTV algorithm with \( \ell_1 \) and \( \ell_2 \) objectives are referred to as 2DFBCS-MTV-\( \ell_1 \) and 2DFBCS-MTV-\( \ell_2 \), respectively.

Of course, the objective functions in (5) and (7) are not convex (since \( \Sigma_k \) is not a convex set and \( \{ X \in \mathbb{R}^{N \times L} : \| X \|_2 = 1 \} \) is also not a convex set), thus we have no guarantee that the algorithms find global minima. If the original signal is known to be non-negative, then the algorithm should be initialized with a projection onto \( \mathbb{R}^{N \times L}_+ \) in each iteration.

### 4 Experiments

In this section, we report results of experiments aimed at comparing the performance of 2DFBCS with that of 2DBIHT. As an illustrative example, we consider the original group-sparse image \( X \in \mathbb{R}^{400 \times 100} \) to contain 10 randomly located groups, where each group is composed of 9 consecutive vertical neighbors, with value 10 or -10. These matrices are then normalized to have unit Euclidean norm. The sensing matrix \( A \) is a 200 \times 400 matrix, whose components are sampled from the standard normal distribution. The variance of the white Gaussian noise \( W \in \mathbb{R}^{200 \times 100} \) is 0.01. Finally, the observations \( Y \) are obtained according to (1).

The thresholds of 2DBIHT-\( \ell_1 \) and 2DBIHT-\( \ell_2 \) are set as \( \tau = 1 \) and 1/\( M \), respectively, and the parameters of 2DFBCS-TV-\( \ell_1 \), 2DFBCS-TV-\( \ell_2 \), 2DFBCS-MTV-\( \ell_1 \), and 2DFBCS-MTV-\( \ell_2 \) are hand tuned for the best results. The recovered signals are shown in Figure 1, from which, we can clearly see that the proposed 2DFBCS basically performs better than 2DBIHT. In general, the algorithms with the \( \ell_2 \) barrier outperform those with the \( \ell_1 \) barrier. Specifically, the 2DFBCS-MTV-\( \ell_2 \) shows its superiority over other algorithms, although the 2DFBCS-TV-\( \ell_2 \) is also good at recovering sparse piece-wise images.

### 5 Conclusions

We have proposed 2D fused binary compressive sensing (2DFBCS) to recover 2D sparse piece-wise smooth signals from 1-bit compressive measurements. We have shown that if the original signals are in fact sparse and piece-wise smooth, the proposed method is able to take advantage of the piece-wise smoothness of the original signal, outperforming the 2D version of the previous method binary iterative hard thresholding (2DBIHT), which relies only on sparsity of the original signal. Future work will involve using the technique of detecting sign flips to obtain a robust version of 2DFBCS.

### References


Abstract

Research done to date has reinforced the potential of Electrocardiographic (ECG) signals in biometric applications. Following an “off-the-person” approach for capturing the ECG signal, the continuous, vital and universal nature of this modality offers several advantages to the field of biometrics. However, some challenges still hinder its widespread adoption in operational settings. One such challenge is the fact that it is still not clear which features from the ECG heartbeat waveform contribute to higher inter-subject differences. In this paper, we assess the recognition performance of various heartbeat sub-patterns (in particular the P, QRS, T, PQRS and QRST subsets), comparing them to the full heartbeat waveform approach, which has been widely seen in the literature. The results suggest that most of the differentiating information comes from the QRS complex. Nevertheless, the entire heartbeat waveform performs better than any of the studied sub-patterns, although closely followed by the PQRS pattern.

1 Introduction

Biometric identity recognition systems, which are essential for security applications, determine the identity of an individual by measuring specific body traits or features. Common examples include iris recognition, fingerprint analysis, and face identification, among many others. The electrocardiogram (ECG) has first been proposed as a biometric trait for person identification in [1], being a universal and easily obtained biosignal that is characteristic to each individual, and hard to reproduce or steal without the individual’s knowledge. A review of the current research in ECG biometrics has been recently published by Odinaka et al. [7].

Previous research done by our group on this topic has focused on overcoming the intrusiveness and usability problems associated with the ECG sensor device, which have been a limiting aspect of this modality [2]. We developed a simplified setup, where the ECG is acquired at the ECG sensor device, which have been a limiting aspect of this modality over the past years. It is still not clear which features from the ECG heartbeat waveform contribute to uniquely characterize an individual. We try to address this aspect here. As shown in Figure 1, the ECG signal is composed by a repeating pattern of waves, the P, QRS and T complexes, as a result of the electrochemical activity in the heart necessary to contract its various chambers in the correct order to effectively pump blood. For identity recognition, our approach has been based on the entire set of the PQRST complexes, whose amplitude values constitute the features fed to the classifier. Our focus with this work is to assess the recognition performance of various heartbeat sub-patterns, i.e. subsets of the typical ECG waves. In particular, we studied the performance of each isolated complex (P, QRS, T) and the PQRS and QRST subsets, comparing them to the full PQRST heartbeat waveform. We tested each ECG sub-pattern using the same approach adopted in [9], using either a k-Nearest Neighbors (k-NN) or a Support Vector Machine (SVM) classifier.

The remainder of this paper is organized as follows: Section 2 describes the methodology used for the biometric recognition system, which is evaluated in Section 3; finally, Section 4 concludes the paper.

2 Methods

The main steps of our approach are summarized in Figure 2. These include a preprocessing block (comprising filtering, QRS detection and heartbeat waveform extraction), an outlier removal block (to discard noisy or spurious heartbeats), a pattern extraction step to select the portion of the heartbeat to use, and finally, a classification block.

2.1 Data Acquisition, Preprocessing and Outlier Removal

ECG data was obtained following an “off-the-person” approach, using two dry Ag/AgCl electrodes placed at the fingers, which were connected to a bioPLUX Research wireless device, set up for data acquisition at 1kHz and with a 12-bit resolution. We recorded the signals from 63 subjects (14 male and 49 female), with ages ranging between 18 and 50 years (20.68 ± 2.83), over two sessions separated by a 3-month interval (denoted as T1 and T2). In each session, the subjects were asked to sit for 2 minutes in a resting position.

The raw ECG signals were filtered with a Finite Impulse Response (FIR) bandpass filter with a Hamming window of 300 ms, and cutoff frequencies of 5 – 20 Hz. We then employed the modified Engelse and Zeevaart QRS detection algorithm, described by Lourenço et al. in [3], which enables the segmentation and extraction of the individual PQRST heartbeat waveforms from the filtered ECG signal. Having identified the location of the R peaks, we define the PQRST heartbeat as the signal segment corresponding to 200 ms before and 400 ms after each R peak, amounting to a total of 600 ms. However, ECG acquisition at the fingers produces a signal with lower signal-to-noise ratio, and more prone to noise artifacts, than typical medical-grade setups. Therefore, we apply an outlier removal step, using the DMEAN algorithm described in [4].
rule out noisy heartbeat segments from the decision process, enhancing the global system performance.

2.2 Pattern Extraction

With the goal of exploring which characteristics of the heartbeat waveform contribute to uniquely characterize an individual, we devised a pattern extraction block that selects the portion of the heartbeat waveform to use for classification. In this simple study, the time delimiters of each sub-pattern were defined a priori for all subjects, disregarding possible subject-specific latencies in the ECG heartbeat. We chose this approach due to the fact that identifying the locations of the P and T waves in ECG acquired at the fingers is more challenging than locating the R wave, producing a high number of false positives. As shown in Figure 3, we considered as the P wave the first 120 ms of the heartbeat waveform, between 120ms and 300ms as the QRS complex, and from 300ms onwards as the T wave. Following the selection of the heartbeat sub-pattern, we applied an additional smoothing of the data by computing means of 5 templates.

![Figure 3: Location of the sub-pattern delimiters (dashed vertical lines).](image)

**2.3 Classification**

Our recognition methodology is based on a template matching approach, following the methodology described in [5], which uses the sub-patterns themselves as features. We compare the performance of $k$-NN with $k = 3$, and an SVM with a linear kernel. We devised two experiments to assess the classifier performance: training the classifier with data from session T1 and testing on T2 (T1-T2), and vice-versa (T2-T1). The performance was measured in terms of the two recognition paradigms: identification (the system determines the identity of the user based on the provided ECG) and authentication (the system verifies that the claimed identity of the user corresponds to the one derived from the ECG). For identification, we based the analysis on the Error of Identification (EID), while for authentication we used the Equal Error Rate (EER).

3 Results

The recognition performances (EER and EID) of the various heartbeat sub-patterns studied are shown in Figure 4. We can see that, in every case, the SVM classifier produces better results than the $k$-NN classifier. Regarding the EER (Figure 4(a)), the entire PQRS heartbeat segment (“All”) performs similarly to the PQRS segment, closely followed by the QRS segment. In the T1-T2 experiment, the best result is obtained with the entire segment, while in the T2-T1 experiment the best result is obtained with the PQRS segment, regardless of the classifier used. It is also interesting to note that the QRS segment alone performs almost as good as the entire PQRS segment, whereas the P and T complexes exhibit the worst results when used by themselves. This was expected, given the fact that the position of the P and T waves varies with the heart rate, while the QRS complex remains stable. Further, the results seem to suggest that it is the QRS complex that contributes with most of the information necessary to distinguish between the subjects. The EID (Figure 4(b)) further reinforces the conclusions inferred from the EER, but in this case the best result is the PQRS segment in the T1-T2 experiment, while in the T2-T1 experiment the best results are the entire heartbeat with the SVM classifier and the PQRS segment with the $k$-NN classifier.

![Figure 4: Identity recognition performance for the studied heartbeat sub-patterns; “All” corresponds to the entire PQQRST heartbeat segment.](image)

**4 Conclusions**

In this paper we explored five different ECG sub-patterns in a context of biometric identity recognition, with the goal to better understand which ECG features contribute for the recognition capabilities of the ECG. The results suggest that most of the differentiating information comes from the QRS complex. Still, the entire heartbeat waveform performs better than any of the studied sub-patterns, although closely followed by the PQRS pattern. This implies the existence of further potential subject-specific information in the P and T waves, especially if the heart rate is taken into account. The design of a heart rate normalization method must, therefore, be the focus of future research in the field of ECG biometrics.

**Acknowledgments**

This work was partially funded by Fundação para a Ciência e Tecnologia under grants PTDC/EEI-SII/2312/2012, SFRH/BD/65248/2009 and SFRH/PROTEC/49512/2009.

**References**


Abstract

Cell-cadherin adhesion among epithelial cells is essential in the maintenance of structural mechanical properties of tissues and in further ascertainment of homeostasis. E-Cadherin is a protein that plays a pivotal role in intercellular linkage. Alterations to its conformation or expression are often associated with an increase of the infiltrative and metastatic potential of a tumour. Hereditary Diffuse Gastric Cancer (HDGC) is often associated to alterations in the E-Cadherin gene. Genetic testing is thus proposed for early detection, but in the case of missense mutations, pathogenicity cannot be directly inferred.

In this work a framework is employed that resorts to fluorescence microscopy images with tagged E-Cadherin to spatially characterise its distribution and concentration and ultimately assess for deleteriousness. This strategy relies on the fact that native E-Cadherin is preferentially located at the plasma membrane, at high concentrations, whilst it exists in lower and uniform concentrations within the cell. Deviations to this pattern are thus potentially indicative of E-Cadherin malfunction.

Through the use of a supervised learning classifier strategy, and resorting to the computed spatially encoded intensity features, an accuracy of 83.2% was obtained in the classification of samples in terms of deleteriousness, alongside 88.5% sensitivity and 69.0% specificity.

1 Introduction

Cell-cadherin adhesion is crucial for the maintenance and homeostasis of epithelial tissues. This adhesion is assured by adherens junctions, which also have a role in promoting cell interaction. E-Cadherin is the pivotal component of the adhesion complex, encoded by gene CDH1 [1].

E-cadherin destabilisation, whether through its reduced or inexistent expression or through its abnormal form, leads to loss of cell adhesion and concomitant increase in cell motility, which in turn can give rise to an enhanced potential for cell invasiveness of surrounding tissues, characteristic of cancer. As a matter of fact, as for gastric cancer, E-cadherin is a consequence of tumorigenesis in HDGC. Most diffuse gastric cancer types only begin displaying symptoms once they are already incurable, therefore the earliest search for preventive and diagnostic strategies in an early stage to assess the risk or predisposition for disease development. CDH1 genetic screening is advised for early detection, after which prophylactic gastrectomy can be performed to prevent HDGC development [2].

However, detected germline CDH1 mutations vary in nature, extent and location which means their impact on E-cadherin functionality and potential pathogenicity is distinct. 22.1% of the mutations are missense [2]. These are extremely hard to assess in terms of pathogenicity, being commonly referred to as unclassified sequence variants (USVs), and represent a clinical burden. An extended search has been made to find complementary diagnostic strategies, without the desired success.

The prime and most valuable piece of information to look at is familial clinical data. However, the size of the pedigrees is usually too small to perform proper statistical clinical analysis [3].

In silico approaches have also been explored, which resort to computational tools to predict the impact of the mutation on protein functionality and possible pathogenicity. Nevertheless, these approaches involve a high degree of inference and are applicable for just a limited range of the E-cadherin structure. Accuracy has not exceeded 70% [3].

In vitro approaches provide a straightforward output on E-cadherin behaviour, assessing for instance cell adhesion and invasion capability in cells expressing for various E-cadherin variants [2]. Nevertheless, these approaches are time-consuming and often qualitative.

Fluorescence microscopy is an extremely useful approach that allows detecting, quantifying and mapping E-cadherin within and between cells. In cells expressing wild-type (WT) E-cadherin subject to normal conditions, E-cadherin is located in high concentrations along the membrane and in lower, uniform concentrations within the cell. Deviations to this pattern, whether in terms of location or overall intensity, are highly informative of dysfunction, de-regulation or aberrant expression, which in turn hints at pathogenicity [1].

The aim is to quantify these potential behaviours through the computation of metrics that act as descriptors of anomalous E-cadherin behaviour, which are then used within a supervised learning framework to allow for differentiation between deleterious and non-deleterious E-cadherin variants, in an objective and reproducible manner.

2 Methods and Materials

The data used within this study consisted of a set of fluorescence microscopy images. Each image consisted of CHO cells expressing for a given E-cadherin variant under study, where E-cadherin was tagged by indirect immunofluorescence with Alexa Fluor 488 (green). Acquisitions were made with a Carl Zeiss Apotome Axiovert 200M Fluorescence Microscope, with 40x objectives and fixed illumination.

The dataset in question is comprised by 21 wild-type images and 86 images of E-cadherin variants, originating from 19 distinct mutations, of which 17 are deleterious. These range all domains of the molecule, being either cytoplasmic, extracellular or pro-domain located.

The process of attainment of useful descriptors from the images followed the method described in [4]. The aim is to obtain profiles that describe the distribution and concentration of E-cadherin within the cell and intercellular space. The following steps are involved (Figure 1):

1) Poisson-based Denoising, by use of a Bayesian strategy.
2) Profiling, both in terms of extraction of single cell radial (RD) profiles and in terms of cell-pair inter-nuclear profiles (IN).
3) RD and IN map building, through the concatenation of all profiles extracted from a same image.
4) Geometric compensation, by use of a Bayesian strategy, to compensate for the non-perfect sphericity of nuclei and other profiling algorithm limitations.
5) Computation of mean and variance (var) profiles.

![Figure 1: Pipeline of the applied image processing strategy, as described in 4. Following RD profiling, a similar process is ensued as for after IN profiling. Blue – mean; Red- variance.](image-url)
According to the maximisation of Youden’s index criterion, the two best SVM-based and the best KNN based classifiers are presented, with their associated diagnostic performances and feature number (Table 2).

Table 2 Performance of three of the tested classifiers.

The classifier based on SVM and using a 2nd order polynomial kernel with $c=100$ yielded the best outcome, not only in terms of Youden’s index, but also accuracy and reduced number of features. According to the feature selection routine, 5 features were chosen, in the following order: IN-ConcDC, IN-var-μ, RD-Conc, IN-var-Conc, IN-d05. Table 3 and Table 4 present a more insightful view on the performance of the algorithm.

Table 3 Confusion matrix for SVM classifier with $p=2$, $c=100$. Positive represents the cases in which the classifier considered a certain sample as being deleterious.

Table 4 List of available information present in the dataset. The last column shows the outcome of the SVM classifier for each given mutation, in terms of percentage (and number) of samples classified as positive, i.e. deleterious: D - deleterious. N – non-deleterious

14 of the 17 deleterious mutations were classified correctly for more than 60% of its samples, with 9 actually having all its samples correctly classified. For the 2 non-deleterious mutations, only one (364) presented a low predicted level of deleteriousness.

4 Conclusion

In this paper a framework was suggested to predict deleteriousness of a given E-Cadherin mutation through fluorescence microscopy data. A classifier with 83.2% accuracy was obtained, resorting to an SVM formulation and based on feature selection strategy. Furthermore, it was shown that the maximisation of the overall accuracy via 10 fold cross validation. Before feature selection an ordering step was applied, such that the most statistically significant features in terms of the Mann-Whitney U-test came first and were preferentially kept.

The classifier, with the corresponding feature set, was then trained and evaluated through cross-validation, in such a manner that at each cross-validation stage either the whole set of samples of a given mutation was removed or a single wild-type sample was removed.

Given the uneven class sizes of the dataset, the ultimate criterion for the choice of the best classifier was not overall accuracy but the Youden’s index (J)

$$J = \text{sensitivity} + \text{specificity} - 1,$$

where sensitivity and specificity arise from the cross-validation strategy outcome.

3 Results

According to the maximisation of Youden’s index criterion, the two best SVM-based and the best KNN based classifiers are presented, with their associated diagnostic performances and feature number (Table 2).

Table 2 Performance of three of the tested classifiers.

The classifier on SVM and using a 2nd order polynomial kernel with $c=100$ yielded the best outcome, not only in terms of Youden’s index, but also accuracy and reduced number of features. According to the feature selection routine, 5 features were chosen, in the following order: IN-ConcDC, IN-var-μ, RD-Conc, IN-var-Conc, IN-d05. Table 3 and Table 4 present a more insightful view on the performance of the algorithm.

Table 3 Confusion matrix for SVM classifier with $p=2$, $c=100$. Positive represents the cases in which the classifier considered a certain sample as being deleterious.

Table 4 List of available information present in the dataset. The last column shows the outcome of the SVM classifier for each given mutation, in terms of percentage (and number) of samples classified as positive, i.e. deleterious: D - deleterious. N – non-deleterious

14 of the 17 deleterious mutations were classified correctly for more than 60% of its samples, with 9 actually having all its samples correctly classified. For the 2 non-deleterious mutations, only one (364) presented a low predicted level of deleteriousness.

4 Conclusion

In this paper a framework was suggested to predict deleteriousness of a given E-Cadherin mutation through fluorescence microscopy data. A classifier with 83.2% accuracy was obtained, resorting to an SVM formulation and based on feature selection mainly for E-cadherin intensity at the membrane as well as peak compaction and overall variance.

Further studies should consider building a set of classifiers, each accounting for mutations in distinct E-Cadherin domains, since these may lead to distinct behavioural changes from WT.

References


Interactive Air Traffic Control automation in oceanic airspace

Francisco Freitas
Miguel Barão
Rodrigo Ventura

1 Institute for Systems and Robotics
Instituto Superior Técnico
Lisbon, Portugal
2 INESC-ID
Lisbon, Portugal

Abstract

Air traffic controllers’ workload limits impose upper bounds to the amount of traffic manageable in a given air sector for a given time frame. Air Traffic Control (ATC) automation methods open the possibility of reducing this workload by shifting to the machine the tasks of (1) detecting potential conflicts, and of (2) proposing to the controller ATC instructions that prevent such conflicts. We propose a decision support system based on a combinatorial optimization approach using a branch-and-bound method. Given a known traffic situation, we proceed by simulating the trajectories of traffic, taking into account possible instructions to separate traffic. In this study we considered only flight level change instructions, given at report fixes. The cost function employed includes both a measure of vertical deviation from the filed flight plan (FPL) and the total amount of ATC instructions. The multi-criteria problem is solved interactively, as the operator directs the algorithm towards the solution, indicating its preferences at intermediate points in the simulation. As a case study, we analyse the problem of oceanic airspace, where conventional ATC is used due to the lack of radar coverage.

1 Introduction

The current ATC system is already near its full capacity in some of the most busy sectors, and will eventually be overloaded if air travel grows as predicted [2]. Given that controllers are already working at the top of their capacities and that an increase in their workload would most likely be a threat to system safety, an increase in air traffic can only be safely handled by a future ATC system if solutions are found to reduce their workload.

A commonly studied approach to the problem focuses on redesigning the existing airspace organization, by creating functional airspace blocks designed to simplify air traffic control and by allowing aircrafts to fly direct routes between fixes (Free Route). An important field of research aims at decentralizing the ATC system, by passing some of the controller’s tasks to aircrafts, allowing agents to self-organize themselves and to perform conflict detection and resolution autonomously. A study comparing the performances of centralized and decentralized strategies is presented in [3].

Better suited for short-term application is the approach of developing computerized decision support tools to improve the performance of the current centralized system (a conceptual scheme for such tool is shown in figure 1). Solutions derived from this approach may range from simple tools, that detect short and medium-term conflicts without suggesting any solutions to total automation, in which an automatic tool detects conflicts, searches for an optimal solution and issues the instructions to the aircrafts by means of some form of digital data format, replacing human controllers.

Our work follows this latter paradigm, and lies midway between a passive system and total automation. A breakdown of the developed tool is presented in figure 2.

1.1 Approach outline

The developed program receives as inputs the current situation of a set of aircrafts and their filed FPLs inside a certain flight region, and calculates their long-term predicted trajectories within a certain time window, using a point-mass dynamics model and a simplified Autopilot. A combinatorial search is carried out to analyse each plan within the state-space. Pairwise conflict detection is executed for each individual plan. A Cost Function taking into account the deviation from the filed FPL and the total amount of ATC instructions was created, allowing a branch-and-bound method to be implemented, pruning regions of the solution search and reducing simulation time, assuring optimality. A Monte Carlo simulation is ran as a robustness check to test each calculated plan. A set of particles is created for each aircraft, simulating possible trajectories.

At the end of the simulation, the algorithm returns a global plan, consisting of a set of instructions to be issued to each aircraft, and presents this plan to the human controller. The controller may accept the plan, or request the algorithm to search for other solutions, indicating which cost function criteria to improve. This iterative process continues until the controller accepts one of the proposed plans.

As a case study, we chose to analyse the operation in Oceanic Airspace, where conventional ATC is used due to the lack of radar coverage. Restrictions were created so that aircrafts fly their filed horizontal route at the requested airspeed, and trajectory changes are only allowed in the vertical plane, with instructions being issued only at report fixes. The algorithm is run every time a position report is received from an aircraft, ensuring the current advisory is based on the most updated information available.

2 Trajectory prediction

The method for the prediction of the future trajectory of an aircraft is based on the one presented by Glover and Lygeros in [1]. Having at its disposal the current state of an aircraft – position, speed and attitude – its future intention – in FPL format – and an aircraft-specific model – loaded from Eurocontrol’s Base of Aircraft Data (BADA) – the program is able to predict the trajectory of that aircraft using a point-mass model.

A six-state control system is implemented, with its states being: geographical position (Latitude \( \phi \) and Longitude \( \lambda \)), altitude (\( h \)), true airspeed (\( TAS \)), heading angle (\( \psi \)) and mass (\( m \)). The aircrafts are assumed to have three control inputs: engine thrust (\( T \)), bank angle (\( \phi \)) and flight path angle (\( \gamma \)).

Being expressed in a cartesian orthogonal frame, the model presented in [1] is unsuited to deal with large length trajectories, as it does not account for the curvature of the Earth. This limitation was overcome with...
quaternion algebra to calculate each successive position of an aircraft by applying a 3D rotation about an axis containing the center of the Earth to the aircraft position at the last time step. This way, the aircraft horizontal position ceases to be expressed as a \((x, y)\) pair, but rather as geographic Latitude \((\phi)\) and Longitude \((\lambda)\).

The three control inputs form a basic Autopilot, ensuring the aircraft follows the trajectory and airspeed required by its FPL. These control inputs result from proportional-integral controllers for the Thrust and the Flight Path Angle in Climb/Descent phase, and from proportional controllers for the Bank Angle and the Flight Path Angle in Cruise phase.

This continuous-time system is discretized using a fourth-order Runge-Kutta method to solve the ordinary differential equations. From now on, system variables are indexed with a discrete index \(k\), \(i.e., t = kT\), where \(T\) is the integration time step.

3 Conflict detection and resolution

3.1 Cost Function

Several different criteria may be used to evaluate a plan issued to an individual flight – trajectory change relative to FPL, number of instructions issued, total time inside FIR, fuel consumption – and there are also criteria suited to qualify a global plan – how much certain flights are penalized relative to others, controller’s workload peak at a given period.

In this work, two criteria are used in the Cost Function: (1) vertical deviation from the filed FPL (denoted \(D\)) and (2) total amount of ATC instructions (denoted \(N\)). The cost function is obtained by multiplying both criteria by weight coefficients (denoted \(\lambda_N\) and \(\lambda_D\)) and then summing them:

\[
f = \lambda_N N + \lambda_D D
\]  

(1)

3.2 Combinatorial Optimization

The algorithm runs a branch-and-bound method to reduce computation time, assuring optimality. Starting at Current Time (the real time being observed by the human controller), the search advances progressively through time, generating predicted trajectories for each aircraft. Every time an aircraft reaches a waypoint, a node is created and several branches are considered, corresponding to all possible flight levels that aircraft can be assigned to. A lower bound of the cost is computed for each branch, measuring the minimum cost each one will add to the total cost function. The branch with the lowest heuristic is chosen and the corresponding action is executed.

At each node, conflict detection is executed in the time window between the previous node and the current one. If a conflict between a pair of aircraft A and B is detected, the algorithm backjumps to the last node (the most advanced in time) where a decision was made involving either A or B, discarding every node in between, and selects a different branch to proceed the search. If the conflict is resolved, the search continues. If not, the algorithm backjumps again to a node involving A or B, repeating the process until resolution is achieved. A solution plan is found when the simulation time window limit is reached, or when every flight in the scenario has already abandoned the airspace being controlled. After the first solution plan has been found, an upper bound is available and may be used to prune branches in the middle of the search tree, whenever a node has a cost higher than the current upper bound. This greatly reduces the run time of the algorithm, requiring the expansion of fewer nodes. The simulation is concluded when the whole search tree has been explored. Following the branch-and-bound method, branches are pruned whenever their lower bound exceeds the current upper bound (i.e., the minimum cost among all solutions found so far). This plan is selected to be presented as a suggestion to the controller.

3.3 Interactive decision-making

The choice of a two criteria cost function deters the use of the conventional optimization techniques commonly used for single-variable problems. When more than one objective is considered, a multi-criteria decision making problem must be solved. In multi-criteria problems, the concept of solution optimality is replaced by those of efficiency and Pareto optimality. In general, there is not a single optimal solution for a given problem, but rather a set of solutions that are Pareto optimal. This set of solutions may be called Pareto front or efficient set.

In this work, an interactive approach is chosen to solve the multi-criteria problem. The algorithm searches for efficient solutions one at a time, using a fixed ratio \(\frac{1}{100}\). Each time it finds an efficient solution, the algorithm presents it to the controller. The controller decides whether he accepts the proposed plan or he requests the algorithm to proceed the search. In the latter case, he must inform which criteria to improve. The cycle ends when the controller accepts one of the proposed plans.

4 Results

The algorithm was tested for a time window of 4 hours with traffic scenarios generated randomly. From a pool of nominal horizontal trajectories based on real commercial flight plans, a certain number of flights \(NF\) is created and placed randomly at different flight levels and entering the controlled airspace at different times. The chosen controlled airspace was Santa Maria FIR, Portugal’s oceanic airspace. The algorithm’s ability to calculate optimal solutions was verified for the tested traffic densities. To test the algorithm computational performance, running time \(t_{run}\) was measured as a function of \(NF\).

Figure 3 shows total running time \(t_{run}\) to increase in an approximately exponential manner as the scenario number of flights \(NF\) is increased. This clearly indicates the complexity of a given scenario – \(i.e.,\) the computational effort it requires from the algorithm – grows much faster than the state-space dimension, which increases linearly with the number of flights \(NF\). This may attributed to the combinatorial effect as more aircraft are added to a scenario, increasing the number of conflicting aircraft, and forcing the algorithm to explore a much larger percentage of the search tree.

![Figure 3: Running time as a function of NF](image)

5 Conclusions

The developed program has proved its capacity in solving the scenarios that it was subjected to. More exhaustive tests are being conducted, especially focusing on the gain achieved by requesting the algorithm to be optimal, \(i.e.,\) on how much the cost function improves as a consequence of the algorithm not ‘settling’ for the first leaf node of the search tree (know as the greedy solution) and rather exploring the whole tree to guarantee the optimal solution is found.

References


Development of a System for Automatic Detection of Air Embolism Using a Precordial Doppler

Ana Tedim¹
Pedro Amorim²
Ana Castro³

Abstract
Venous air embolism (VAE) is the air bubble accumulation in the right side of the heart. Changes in Doppler heart sound (DHS) are characteristic of VAE, and the anesthesiologist has to pay attention to this event, what is not always possible. This work aims to study different features of heart sound through precordial Doppler that may provide useful information on VAE episodes.

A clinical protocol was designed and DHS was collected at baseline and following infusions of saline with 4 distinct volumes (1ml, 5ml, 8ml and 10ml) and two infusion rates (slow and fast), were given by central and peripheral catheters. Signal pre-processing and the envelope of each signal was extracted, and five features were implemented and evaluated: frequency corresponding to 95% of Welch periodogram power (f95), frequency corresponding to 50% of Welch periodogram power (f50), frequency corresponding to maximum power spectral density (fm), entropy (E) and frequency corresponding to maximum energy of a wavelet transform (ffreqwav). Relation between extracted features and saline infusions were studied and compared to baseline values. A Graphical User Interface (GUI) with a database of Doppler heart sounds and annotations was also developed.

Although features present a high variability, E presents a better performance showing an increase in response to saline injections. The feature with worse results was fm since almost no difference was observed during turbulence episodes. Higher infusion volumes administered by central catheter, resulted in more pronounced responses of the evaluated features.

1 Introduction
Venous air embolism (VAE) is a serious complication that may occur during neurosurgical procedures. It is the result of air bubble accumulation in the heart, causing over distension of the right side and when not detected in time, causes many health problems like cyanosis, or cardiovascular collapse [1, 2].

The accumulation of air bubbles is caused by the existence of pressure gradients between the right side of the heart and the incisional area [1].

To detect and prevent these problems, monitoring protocols are implemented, including a precordial ultrasonic Doppler probe [1] or other detection method like transoesophageal echo. The Doppler probe allows a real-time monitoring of the heart sound [3]. Changes in heart tones, usually referred to as “mill-wheel murmurs” are characteristic of VAE and the anesthesiologist has an important role in the detection of these characteristic sounds, and subsequent intervention [1].

Since the anesthesiologist needs to focus his attention on other tasks, detection of VAE may be disregarded; especially with the occurrence of false positives.

In this context, it is imperative to assist the decision of the anesthesiologist by studying and developing other robust methods of VAE detection. The objective of this work was to study the Doppler heart sound (DHS) under different turbulence conditions.

2 Materials and Methods
To perform data acquisition, a clinical protocol was designed and submitted for approval to the Department of Education, Training and Research, and to the Ethics Committee of Santo António’s Hospital, Centro Hospitalar do Porto. After study approval, data collection was initiated.

During intraoperative management, the precordial Doppler was placed on the patient’s chest according to manufacturer recommendations [2, 4, 5]. All patients were submitted to a total intravenous anesthesia (TIVA) of propofol and remifentanil [6]. To study which features allowed a better discrimination between normal blood flow and turbulent blood flow, the designed clinical protocol included a series of saline injections through peripheral and central catheters, at different rates and volumes of infusion, as way of study how these conditions affect changes in Doppler sound. The protocol was designed in this way because injection of air bubbles was not viable, and the occurrence of VAE is low for the available time span. The administrations of saline started after the patient was stabilized following induction. In total, each patient received 16 saline injections (with intervals of 15 seconds) with volumes of 1ml, 5ml, 8ml, and 10ml at two distinct administration rates (one as slowly as possible and the other faster) by peripheral and central catheters.

Data was collected using the RugloopII Waves. Besides data from the monitors, demographic data of the patients was registered. DHS was collected at 8kHz, and all injections moments were registered by the researcher for the following analysis. Matlab 2012 was used for signal processing.

2.1 Heart Sound Segmentation
After signals acquisition, these were downsampled to 2kHz and normalized. Frame of 3 seconds corresponding to DHS were analyzed, since this is the adequate time to detect at least one complete cardiac cycle [7].

Pre-processing of the signal included filtering the unwanted frequency components. A Butterworth low pass filter was applied to the DHS with a cut off frequency of 1800Hz in both directions to avoid phase distortion. The Savitzky-Golay filter [8] was also implemented, because it smoothes the DHS [9]. To apply this filter to the DHS, a polynomial of order 3 was chosen since the Doppler heart sound has the shape of a sinusoid. The chosen frame size was 21 because higher frame sizes highly distort the signal. These filters were applied to the DHS with different purposes: the first to remove frequency components outside the band of interest, and later extract signal features for analysis; and the second to provide a smoothed version of the signal, and construct the envelope for heart cycle identification.

Following signal filtering the next step was to extract the signal envelope and segment the signal into heart cycles. The energy of the signal was calculated using the Savitzky-Golay filtered DHS in windows of 0.5 s. The size of window chosen corresponds to the size of one cardiac event. Then peaks of the envelope were detected using the function “findpeaks” with the characteristic “minimum peak distance” this last with the value of 0.1 s. The detected peaks were then interpolated with a cubic Hermite spline to provide a smooth envelope of the signal.

After obtaining the envelope, the cross-correlation was calculated for two consecutive frames. After this step, local maxims of the cross-correlation were detected, and the heart rate was identified.

2.2 Doppler Heart Sound Features
Frequency content of the DHS was analyzed by estimating the PSD of the signal through the Welch method. The PSD Welch was obtained using a 150 points window with 50% overlap, and a Blackman window.

After obtaining the PSD for each heart cycle, three features were extracted: the frequency corresponding to the limit of 95% of PSD power (f95), the frequency corresponding to 50% of the PSD power (f50) and the frequency corresponding to the maximum power peak of the PSD (fm). The method was applied to the signal using a moving window of 0.1 seconds and the vectors stored.
Continuous wavelet transform (CWT) analysis was also used to study the DHS. The scales used were between 1 and 10 at distance of 0.1. In this case, the chosen wavelet was the "Mexican Hat Wavelet" once it was demonstrated to provide good results in the analysis of the Doppler sound [10]. CWT was applied to intervals of 0.1 seconds of the signal to extract the respective scales. With the coefficients obtained, the points of maximum energy of the signal and the respective scales were extracted. Using the function called “scal2freq” that correlates the frequencies and the scales, it was possible to know in which frequency band the energy was higher (freqwav) and store its the values in a vector. Again, a vector with the average of this feature for the corresponding heartbeats of two seconds of each injection was created.

An additional feature proposed in this study was the Shannon Entropy (E). This feature was also implemented at each segment of 0.1 seconds and the values with the mean of this feature for the heartbeats of the two seconds after each infusion was stored in a vector for future evaluation.

2.3 Features Evaluation and Graphical User Interface

After extracting all features and store them in a text file, these were exported to Excel (Office 2010) for the remaining processing. The features average value for each injection and patient were calculated. Also the graphical representation that relates the volume of infusion with the ratio injection/baseline changes for different administration rates and catheter were performed for all features.

To allow automatic extraction of the features from each signal, a graphical user interface (GUI) was developed. The GUI allows the user to choose the signal to analyze, displaying patient’s demographic information, estimated heart rate, and implemented features baseline.

3 Results and Discussion

6 patients were enrolled in this study but only 4 acquisitions were afterwards studied because an error occur during the infusions of saline. All acquisitions were made in females (means: age=73 years; height=159 cm; heart rate=75 beats/min; weight=62 kg).

The creation of an interface was useful to help the user to see the features extracted and correlate them with the original signal.

3.1 Features Evaluation

The ratio between the baseline features and features following injections was related with the infusion volumes in terms of rates and via, for each feature as demonstrated in Figure 1, for E feature.

E shows to be the feature with best results because almost all values corresponding to the infusions are higher than the basal value, which mean that this feature detected almost all sound interferences caused by the entry of saline infusion via central or peripheral catheters.

4 Conclusion

In order to improve the detection of VAE episodes using the precordial Doppler, the first study in humans was proposed.

Pre-processing of the signal to eliminate out of band noise and normalization was implemented; a new method for signal envelope retrieval, and heart rate estimation, based on a Savitzky-Golay smoothed DHS was applied with success to DHS segmentation; finally, detected DHS events were analyzed for each patient, with several time-frequency features extraction for analysis.

Five features were proposed in this study and evaluated in response to saline injections. There was a high variability in the results because the number of patients was low. The feature with best results was the entropy since almost all saline infusions caused changes in Doppler sound, detectable with this feature, when compared with baseline DHS. The feature with worse results was Welch periodogram maximum frequency, once almost no changes were observed.

In a general way, it is possible to conclude that higher volumes of infusions turns the changes in Doppler sound more evident and detectable when using these features. Also the detection of turbulence is more evident when the infusions of saline are administered by central catheter, probably because the infusion enters directly in the heart, where the Doppler probe is placed (externally).

Future work will include the increase of the number of acquisitions to improve the database created, the acquisition and study of heart sounds when VAE events occur to evaluate the robustness of the features in detecting air in the circulation, and the adaptation of the GUI for online analysis. This would allow to implement an advisory system that could aid the anesthesiologist detect VAE episodes more rapidly.

Acknowledgments

Work partially funded by the project I-CITY - ICT for Future Health NORTE-07-0124-FEDER-000064.

References

Abstract

Determining the power generated by wind turbines at future times is important for unit commitment planning and maintenance scheduling. Wind speed must be predicted to estimate wind power generation capacity. Because wind power is weather dependent, and therefore, is variable and intermittent over various time-scales. So, an accurate forecasting of wind power is recognized as a major contribution for reliable large-scale wind power integration taking profit of economics gains. This paper explores a new approach using the by fuzzy clustering algorithms, for obtaining the one day characteristics curves of wind power, with an Artificial Neural Networks (ANN) to forecast the average hourly wind power, for an ahead horizon of 24 hours.

1 Introduction

A rapid expansion of wind energy [1] and [2] has led to new challenges in turbine control, plant operations, production planning, condition monitoring, and maintenance. Advances in research over the past years have provided numerous solutions to different problems. The research has focused on variability of wind speed, increase of power generation efficiency, and reduction of the generation cost. This research effort is the result of difficulty into predict the wind power production value, as well as its maximal or minimal values and their occurrence instants.

The increased incidence of wind power in the energy network causes an increase of the unpredictability factor of energy production. The cause of this problem is that the wind velocity and its orientation are considered as one of the most difficult meteorological parameters to forecast. The forecasting technique employed depends on the available information and the time scale in question.

The intermittency regime of wind speed and its direction are strong factors that determine the extent to which energy is generated from the wind turbine. This problem is exacerbated by the fact that wind energy cannot be stored and cannot be easily ramped up to meet load requirements [2]. To address these problems, wind power forecasting methods can be used to plan unit commitment, scheduling and dispatch by system operators, and maximize profit by electricity traders. Because wind power is weather dependent, it is variable and intermittent over various time-scales. A good forecast of the produced power is, therefore, very important, so that an accurate forecasting of wind power is recognized as a major contribution for reliable large-scale wind power integration in the electric networks (of producers and consumers). This demand of prediction accuracy motivates researchers to propose accurate short-term forecasting models of wind power.

The wind power forecast should be based on the actual wind velocity forecast or on the output power of the wind turbines. Several mathematical models which hybridize weather forecasting models and statistical techniques have been proposed in the literature [3-4] to forecasting wind speed values. Also, in many cases, these systems use statistical down-scaling processes including auto-regressive models [5], artificial neural networks [6] or support vector machines [7], as a final step to improve the wind speed forecasting of the system.

However, the wind power can be a more viable prediction parameter than wind speed for power generation purposes. This premise is supported on the fact that predicting wind speed can be converted into power output using power curves or in the following equation which relates the wind turbine’s power output to wind speed:

\[ P = \frac{1}{2} C_p \rho A_v v_w^3, \]  

where \( C_p \) is the coefficient of performance, \( \rho \) is the air density, \( A_v \) is the area swept by the blade and \( v_w \) is the wind speed at right angles to the turbine’s blades-face.

The wind power signal is responsive to the oscillation and intermittence factors that happen with the low speed wind regime which results in a power output signal with better quality, less noisy corrupted and so more predictable. By using the power curves and equations in converting wind speed to wind turbine power output can be made through aggregate forecasting [8]. For these reasons, in this paper we are focused in task of wind power forecasting.

This work proposes a new kind of short-term forecasting model, based on a hierarchical structure, which combines distinct models. These are results of an identification process that, from analysis of historical time series of potential of electric wind power \( \{P_i^v\} \), finds daily pattern sequences of this data series. The normalized values of \( P \) are calculated from the collected wind speed from a meteorological station and used to training and validation of the ANN model. The values of wind power from the previous hours are grouped into clusters according to their similarity together with an ANN that uses each of the clusters as an input to forecast the power production of wind farm one day ahead. A new learning method for a short-term forecasting of the wind power, by using a ANN model and clustering approach, is proposed to predict the electric power production into an aeronial farm.

The test of this new forecast scheme, jointly with the proposed identification method, has shown excellent results in the task of short term wind power prediction. Thus, it is an alternative and effective method that can help to predict in real-time the wind energy produced, allowing to make a good planning and managing of the balance between consumption and production over the power grid.

2 The forecasting model

The proposed approach to forecast wind power generation is based on a wind speed prediction. Previously, all data of wind speed are converted to wind power, and the latter, normalized.

This model is divided into three phases: initially, the data values of the wind power of each 24 hours of a day are grouped into several clusters in an exclusive form based on the degree of similarity between the different data, grouping the values according to their typical characteristics (clustering). Next, each data vector (a day of power wind values) is decomposed by projection on the basis vector in the set of prototype clusters centers. As final step, the relationship between the projections values and the membership values to the clusters is modelled by an ANN. This forecast model uses as inputs the wind power values and the cluster that most identifies with the class of winds power to predict. With these predicted values of wind power we can easily forecast the wind power production.

In this paper the feature vector represent the daily wind power \( x_h = \left[ P_1^w, \ldots, P_h^w \right] \) with 24 hourly normalized wind power values. The sub-script \( h \) is number of day while the upper scrip \( h \) is the hour of day.

Data clustering algorithms is used to divide data elements \( X = \{x_1, \ldots, x_o\} \), with \( x_i \in \mathbb{R}^n \), into \( c \) classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible. With fuzzy c-means, the centroid of a cluster is the mean of all points, weighted by their degree of belonging to the cluster. i.e., \( X_c = \{x_1, \ldots, x_o\} \) is a c-tuple objects prototypes. In this paper, these prototypes clusters are the characteristics curves which can be used as base vector decompositions of daily wind power vectors. The matrix \( \mathcal{U} = [u_{1}, \ldots, u_{c}] \) is the fuzzy partition, \( u_{ik} \in [0,1] \), representing the membership degree of feature point \( x_h \) to cluster \( i \), satisfying the following constraint \( \sum_{i=1}^{c} u_{ik} = 1 \).

From this methodology of analysis, the daily wind speed \( x_h \) is related by the membership vector \( u_h = [u_{1}, \ldots, u_{c} \ldots, u_{c}] \). Found the relations between the pair wise vector \( u_h \) and the correlations matrix of power wind with prototypes \( X_h \) is the main task for the design of the proposed forecasting model.

The algorithm to predict wind velocity is divided in the following steps:
Step 1: In the first part, the FCM algorithm is applied to the dataset. The aim is to detect patterns in existing data in order to build groups, the most homogenous possible. As a result, we get the centres of each cluster corresponding to a typical curve (curves prototypes). The result of this classification is expressed by the matrix $U$ and the cluster centres, $X_c$.

Step 2: The next step is to use a similarity measure (inner product) to establish the relationship between the centres of the clusters, $X_c$, and the power wind curves in order to classify the patterns of prediction. The result of this classification is expressed by the matrix $V$.

Step 3: The next phase establishes the relationship between the matrix $V$ and the matrix $U$ with the membership degrees of each curve with respect to the cluster centres. This relation is realized by a neural networks, $v_{ij} = ANN(u_{ik})$ and their weights are adjusted by the backpropagation algorithm, with $v_{ij} \in V$ and $u_{ij} \in U$.

Step 4: Output of the model: computation the next day forecasting wind curve $Y$, given by:

$$Y = X_c \times V \quad \text{with} \quad V = ANN(U).$$

3 Results

In this work we make the prediction of wind speed from historical data available from National System of Water Resources (SNIRH) (see http://snirh.pvt.fr/) from a set of station near of locations of twenty wind farms. The wind power prediction model was building using the normalized data set from the years: 2007, 2008 and 2009. The wind speed data is available for every hour and grouped in a vector of 24 hours of speed wind. The proposed algorithm use these data to building a predictive model, which is part of the clusters centers (mains typical wind power curves) and a ANN which related the clusters membership values, $U$, with decomposition parameters, $V$.

In the context of this application example, the clusters centers $X_c$ are typical daily wind power curves, represented in Figure 1 by dotted red points. For each of the 20 clusters, the curves with membership degree greater than 0.5, are also depicted.

Normalized numerical forecasting results with the proposed approach are shown in Figure 2, for the first and last 1000 hours of validation test (i.e., some continuous days). Here the blue line is the real wind power and red line their forecasted curve by the model. The forecasted wind power curve follows quite well the real curve with a RMS error of 0.0031 and 0.0038 for, respectively, the training and test data. These lower prediction errors, for the train and test periods, show that the proposed model have a good performance, known the values of composition of the wind day patterns (i.e., the memberships values on clusters).

4 Conclusions

In order to increase the penetration of wind generators to the electrical grid, proper management of the dispatch of the electrical system must be acquired. In order for this to occur, it is important to have reliable and accurate techniques to forecast the wind speed in the short and middle term. It will be very useful to have predictions in advance for a full day, combining the weather forecasts with specific topological locations of wind farms.

In this paper we have presented a model for daily (24 hours) wind power forecasting. The proposed approach combines clustering techniques and a classification neural network applied to wind speed estimation. The results from the prediction of a wind farm in Portugal show that the proposed approach provides accurate and efficient wind power forecasting. The proposed NN-Fuzzy Clustering method was shown be a robust and accurate forecast model. Models that gave predictions with errors in the range below of 8%, for 365 days test. When the output of these models was compared to the actual data, the results proved to be very good.

References

Prediction of solar radiation using artificial neural networks

João Faceira
joao_faceira14@hotmail.com
Paulo Salgado
psal@ual.pt
Departamento de Engenharias - ECT, Universidade de Trás-os-Montes e Alto Douro, Portugal
Departamento de Engenharias - ECT, Universidade de Trás-os-Montes e Alto Douro, Portugal

Abstract
The use of technologies that allow the prediction of weather for a particular location is very important for the everyday man and several areas of science (agriculture, electricity production). A stochastic model for simulating global solar radiation is usefully in power systems reliability calculations, which importance is increased by the use of photovoltaic power generation. The main objective of this paper is to present an algorithm to predict hourly solar radiation in the short term, combining information’s about cloud coverage level and from solar radiation historical registers, which increased the performance and the accuracy of the forecasting model. In this paper is proposed a new and efficient method capable of forecasting 1-day ahead solar radiation during cloudy days by using an artificial neural networks.

1 Introduction
Hourly solar radiation data forecasting has significant consequences in most solar applications such as energy system sizing and meteorological estimation. Accurate forecasting of solar radiation improves the efficiency of the outputs of many applications.

The global solar radiation shows not only regular yearly and daily variations but also a random behavior. The yearly and daily variations can be described in a deterministic way while the random behavior has a high correlation with the state of the atmosphere. Outside the atmosphere the solar radiation (OASR) can accurately be determined [1] and the atmosphere will induce the randomness [2]. The transmittivity of solar radiation in the atmosphere depends on various factors, e.g. humidity, air pressure and cloud type, but where factor that has a great impact on the transmittivity is the cloud coverage. By assuming a deterministic relation between cloud coverage and hourly global solar radiation, the need for measurement of the latter disappears.

Classically, the solar radiation data can be regarded as a random time series produced by a stochastic process, and its prediction depends on accurate mathematical modeling of the underlying stochastic process. Using an accurate model, the prediction is mathematically the conditional expectation of the data given the generally nonlinear model, from present relevant information and the past data samples (historical behavior). On the other hand, the computation of such conditional expectation requires the knowledge of the distribution of the samples including higher order statistics. Since the available or recorded data is finite, such distributions can be estimated or fit into pre-set stochastic models such as Auto-Regressive (AR) [3], Auto-Regressive Moving-Average (ARMA) or Markov [4]. Another new approach, it is the use of ANNs [5-7] that is trained to predict results from available historical examples, deal with nonlinear modeling building from a train processing that extract relationships from data available. This model can be used for forecasting or modeling if we guarantee that the structure of the ANN is adequate (in number, types and links of elements - Neurons) and the training process is sufficiently efficient to extract the relevant relationships from rich data set. By combining the solar radiation model with an ANN model of cloud coverage the forecasting method could be even more suitable.

The level of cloudiness or transmittivity is expressed in Oktas [8]. An okta is a unit of measurement used to describe the amount of cloud cover at any given location such as a weather station. Sky conditions are estimated in terms of how many eighths of the sky are covered in cloud, ranging from 0 oktas (completely clear sky) through to 8 oktas (completely overcast). By using cloud observations as input for the simulations, the local solar radiation can be estimated with accuracy. Also, with a good cloudiness forecasting values, from the medium term to the long run horizon, the prediction of solar radiation can be forecasted. This information is currently provided by official sources of weather forecasting, meteorological satellites and ground weather stations [9]. Here, the proposed model will go adjust the cloud coverage information taking in account geographic specificities of the solar farm as well as the recorded historical weather data.

In this paper a new low complexity approach is proposed to predict solar radiation for a period of one ahead day based on the knowledge of cloudiness forecast provided for the day. With this solution we hope to increase the quality of management of the solar panels systems and so contribute for a best electric energy efficiency production. This can be obtained by including a stochastic model for the short-term variations and simple model has been proposed. Errors and limitations of the model are estimated and discussed.

2 NN forecasting model
Artificial neural network (ANN) is an interconnected group of artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation. ANN’s are increasingly nowadays an important role in the simulation of natural intelligence because they are able to learn from examples, deal with incomplete data, deal with a large number of data related to several parameters, are able to focus on essential data and solve nonlinear problems. Once properly implemented and trained the ANNs are able to make predictions of several problems, among other things.

In the figure 1 is represented the structure of the forecasting model, where the inputs are the expected level of cloudiness, the date and time, the geographic localization and as the output the solar radiation forecasting values.

![Figure 1: Structure of the forecasting ANN Model.](image)

3 Development
This section will present the main steps undertaken for the development of the algorithm in *matlab* for prediction of solar radiation over two years in the region of Viana do Castelo. In a first step, the analysis of hourly solar radiation does not take into account the presence of clouds in the atmosphere. In a second step we proceeded to create a *matlab* script that implements a neural network that predicts hourly radiation incident given the cloudiness. Data collected in the Weather Station located at Ponto da Barca [9], Viana do Castelo, was used for test and validation of the forecasting solar radiation model. These data are from the years 2009 and 2010 and they have the format of the Table 1.

<table>
<thead>
<tr>
<th>Date, Time, Latitude, Longitude</th>
<th>OASR</th>
<th>Solar radiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-01-2010 09:00</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 10:00</td>
<td>134</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 11:00</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 12:00</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 13:00</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 14:00</td>
<td>135</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 15:00</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 16:00</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 17:00</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 18:00</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>01-01-2010 19:00</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Radiation data between 9am to 9pm on January 1, 2010.
4 Tests and Results

In an initial step was not foreseen solar radiation taking into account the absorption factor which is mainly due to the passage of clouds in the atmosphere. In Figures 2, 3 and 4 are plotted the real solar radiation and prediction values from OASR model.

**Figure 2:** Prediction of radiation without cloudiness.

**Figure 3:** Forecasting of solar radiation for five days without cloudiness.

**Figure 4:** Forecasting of solar radiation for 5 days with cloudiness.

In Figure 3 we can see that the actual radiation is very close to the theoretical/predictable radiation. However, if we make an analysis for another set of days, as shown in Figure 4, we notice that the actual radiation is no longer close to the expected radiation. Such evidence can be explained by the passage of clouds that can absorb solar radiation. To solve this problem an ANN was implemented to predict the cloudiness, from the information done by the National Weather forecasting institute for the local, quantified in 5 discrete levels of cloudiness and for 4 periods of the next day. With the ANN - OASR joint model, the new results of solar radiation forecasting can be seen in figures 5, 6 and 7. The RMS error of forecasting was 0.031 and 0.038, respectively for the training and validation data. In view of these results, we can conclude that forecasting values of solar radiation are very close to the measured values. The extrapolation of these results can be used to the prediction of power production by a local farm of photovoltaic panels.

5 Conclusion

In this work, a new model for daily solar radiation forecasting, is comprehensively evaluated through the combination of the OASR model with the ANN prediction model. This forecasting model predicts solar radiation for 24 hours, for the next day. The values obtained with the use of neural networks (tuned by historical data values) show that the tools developed allow to estimate the values of radiation with great precision. With this information, the power output of a photovoltaic production center can be estimated and so allowing the implementation of an efficient strategies of managing energy.

**Figure 5:** Prediction of radiation with clouds.

**Figure 6:** Prediction of radiation with clouds.

**Figure 7:** Prediction of radiation with clouds.

References


Large Scale Automatic Detection of Sub-km Craters Using Texture Information

Marlene Machado¹
Lourenço Bandeira¹
Jorge Salvador Marques²
Pedro Pina³

¹CERENA, Instituto Superior Técnico, Lisbon University
²ISR, Instituto Superior Técnico, Lisbon University

Abstract
Automatic detection of craters is one of the most studied subjects in the application of automated methods to the analysis of planetary surfaces, since these features exist in all terrestrial planets in a wide dimensional range. Their densities, patterns and morphologies can provide valuable data on the age and evolution of a landscape. The focus of this research is the study of a sub-km crater detection algorithm applied to a large region covered by a high resolution image of Mars.

1 Introduction
Craters are topographic features on planetary surfaces resulting from impacts of meteoroids. They are found on all hard-surface bodies in the solar system but are most abundant on bodies such as the Moon, Mars or Mercury where they can accumulate due to slow surface erosion rates. Importance of craters stems from their utility to provide relative chronology of different planetary surfaces [12, 14]; simply put, heavily cratered surfaces are relatively older than less cratered surfaces. Because statistics of crater sizes form the basis for geologic stratigraphy, crater counting is a routine activity in planetary science [5, 6]. Presently, all crater surveys are done by means of visual inspection of images resulting in databases which are either spatially comprehensive but restricted to only the largest craters [1, 4, 7, 11], or size comprehensive but limited to very specific geographical location. The size distribution of craters can be well approximated by the power-law [12]; large craters are rare and small craters are abundant. Counts of large craters must be collected from spatially extended regions in order to accumulate sufficient number of samples for accurate statistics. Thus, geologic stratigraphy based on manually collected databases has coarse spatial resolution. Finer spatial resolution of the stratigraphy can only be obtained from statistics of smaller craters, and the only viable means to obtain spatially comprehensive databases of smaller craters is through automating the process of crater detection. Although crater detection algorithms (CDAs) have been extensively researched (see [3, 10] for the overview), the previous research focused on detecting larger craters from coarse resolution images. Such CDAs are not expected to work well for detection of smaller craters from high resolution images. In this paper we report on the development of CDA for surveying sub-kilometers size craters from high resolution images.

2 Methodology
The purpose of this work is to contribute toward development of a CDA especially designed for surveying sub-kilometers size craters in high resolution planetary images. The key insight is an observation that discriminative features utilized to obtain a set of crater candidates do not have to be the same as discriminative features utilized to extract craters from the set of candidates. Our CDA is particularly well-suited for accurate selection of craters from amongst the candidates using the combination of Haar-like image texture features and a classifier [2]. SVM-Support Vector Machines in the current study. Shape filters are used to identify crater candidates - portions of an image that contain crescent highlight and shadow shapes indicating possible presence of craters [13], and then texture features are calculated for these crater candidates, such as in [8]. The features, collected from a small portion of candidates for which a label (crater or non-crater) is manually determined, are used to train the SVM classifier that is then utilized for classification of all other candidates into craters and non-craters.

3 Dataset
To test the performance of our CDA we used a portion of the Mars Express HRSC nadir panchromatic, 12.5 m/pixel image of Mars (# h0905_0000). The selected scene is centered on Nanedi Valles (4.9°N, 49.0°W); it extends ~75 km (north-south) by ~56 km (east-west) and covers ~4220 km² of Noachian terrain. In this scene we have manually cataloged 6710 craters having diameters between 18 to 6600 meters. The Nanedi Valles passes through the middle of the scene introducing some heterogeneity of the terrain. In order to account for this heterogeneity we have divided the scene into three sections labeled West, Central (containing Nanedi Valles), and East. A total of 332 craters and 664 non-crater candidates were selected from the northern half of the west section to train the classifier. Note that the training set contains only 5% of the craters in the scene and is sampled from approximately 10% of the area of the scene. Thus, our experiment corresponds to a likely use of the CDA, where user wants to train the CDA on small image and use the CDA to find craters in the large image.

4 Results
The objective of our CDA is to automatically survey craters larger than 200 m (16 pixels) but smaller than 1 km (80 pixels) in diameter. The lower limit stems from the minimum number of pixels required by the classifier to make a determination, the upper limit is arbitrary – crater larger than 1 km in diameter on Mars have been already cataloged [9]. Figure 1 shows the results of applying our CDA to the test scene.

Figure 1: Craters detected by our CDA in a 7,000x4,500 pixels (75,000x56,250 m²) test image. West (left), Central (middle) and East (right) regions. True detections are shown by green circles and false detections are shown by outlines. Note that most of the larger craters are outside the detection size range.

To evaluate the performance of our CDA we measured the detection percentage $D = 100 \times TP/(TP + FN)$, the quality percentage $Q = 100 \times TP/(TP + FP + FN)$ and the branching factor $B = FP/TP$. Here, TP stands for the number of true positive detections (detected craters that are actual craters), FP stands for the number of false positive detections (detected craters that are not actual craters).
(detected craters that are not), and FN stands for the number of false negative “detections” (non-detection of real craters). D can be treated as a measure of crater-detection performance, Q as an overall measure of algorithm performance, and B as a measure of delineation performance. Table 1 resumes these results.

Table 1: Performance of our CDA

<table>
<thead>
<tr>
<th></th>
<th>D (%)</th>
<th>Q (%)</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>West section</td>
<td>80.3</td>
<td>64.9</td>
<td>0.29</td>
</tr>
<tr>
<td>Central section</td>
<td>50.8</td>
<td>48.9</td>
<td>0.08</td>
</tr>
<tr>
<td>East section</td>
<td>74.7</td>
<td>66.3</td>
<td>0.17</td>
</tr>
<tr>
<td>Entire test site</td>
<td>72.6</td>
<td>62.3</td>
<td>0.23</td>
</tr>
</tbody>
</table>

5 Conclusions

Our CDA combines the efficiency of shape analysis in identifying crater candidates with precision of texture features/SVM classifier in distinguishing between craters and non-craters. First, the weakest link (from the accuracy point of view) of the present algorithm is the shape-based identification of crater candidates. Of all the identifiable ground truth craters within the size range in question, 252 are not covered by any crater candidate thus becoming false negatives by “default”. With all ground truth craters covered by candidates the detection rate of our algorithm could be improved to up to 79% using the same training set. Thus, further research should first concentrate on improvements to the shape-based identification of the candidates. One way to increase the number of candidates is to relax the shape filter criteria. However, since the utility of the shape-based identification of the candidates is its computational efficiency, this efficiency should not be sacrificed by over-relaxed shape criteria. The training requirements are currently not oversized (only 5% of the craters are utilized for training) and the performance of the new algorithm is already relatively high (TDR = 73%, B = 0.23) for its utilization in research-oriented crater counting application, thus a larger and more representative training dataset must also be built.

References

Abstract

An interactive platform supported on machine learning algorithms to automatically detect impact craters on remotely sensed images from planetary surfaces is presented in this paper. It constitutes the user-friendly interface of previous developed algorithms in order to permit their widespread usage by a larger number of users. It is organized into three main modules: data management, generation of candidates, training and testing. The modularity of the application will permit the future inclusion of additional options and features of interest in an easy fashion.

1 Introduction

Impact craters are bowl-shaped depressions formed by the impact of high-speed meteoroids on a planetary surface. Craters are a very common feature throughout the Solar System, that are intensively used in studies of diverse nature, namely, for chronostratigraphic analysis or for reconstructing past climates [3]. The dimensional range varies from a few meters to some thousands of kilometres in diameter being their perception dependent on the resolution of the sensor used. The largest craters of the more visited bodies by space probes are already catalogued and can be considered complete up to a certain dimension: 1 km for Mars [8], 5 km for the Moon [12] and 10 km for Mercury [4]. The spatial resolutions of the mosaics used for cataloguing the craters on each of these bodies were anyhow of low resolution, namely, above 100m/pixel. Currently, there exists imagery of metric and also centimetric resolution that can be used for large scale mapping, meaning that a myriad of craters with diameters in the range of few tens of metres are waiting to be identified. Consequently, the available databases, which already reach for Mars some few hundreds of thousands of craters, can be hugely expanded. But for achieving that purpose, automated approaches are needed. The last decade has seen the appearance of an increasing number of automated crater detection algorithms (CDA) that have rapidly evolved from more traditional approaches of edge detection or template matching to adaptive and more powerful methods like AdaBoost or Bag-of-Features. For a recent and complete review on CDA consult [9]. Our methodological contributions to the field are already permitting their use in the upgrade of catalogues for Mars [10] and one of its moons, Phobos [11].

The necessity felt to make our algorithms available for a larger number of people, especially for those whose expertise does not fall into a deep knowledge of the methods but in planetary science, has pushed us to integrate the algorithms developed into a platform with a user-friendly interface. This paper addresses the presentation of this application, which is mainly organized around four main modules: data management, the generation of crater candidates in order to reduce the dimension of data to be subsequently analysed, the training of the classifier with a selected type of features of positive and negative examples and finally with the testing phase.

2 Data management module

The preparation and selection of the datasets that are going to be used along the whole procedure is defined within this module.

The selection of training and testing images or the indication of ground-truth datasets (if available) is performed here. Many other operations (image zoom, output file management or text file edition) are also available at this stage. The typical interface menu with text fields and image windows is shown in Figure 1.

3 Candidates generation module

In this first phase of the detection sequence an efficient identification of a set of viable crater candidates is performed in order to significantly reduce the number of examples to be analysed by a classifier for separation of true craters from non-craters. The algorithm behind this procedure was originally proposed by [13] and improved by [1] and assumes that the craters in an image can be recognized as a pair of crescent-like highlight and shadow regions. It is a multi-scale morphological approach using attribute filters [13] for simultaneously detect the characteristic pair of each crater candidate. In this module, the users are asked to choose between different attribute filters and to define the range of crater diameters to search for.

Figure 1: General interface menu

Figure 2: Visualization window for displaying large mosaics and the working grid selected with indication of overlapping between adjacent images.
4 Training module

Craters are detected from amongst the previously selected candidates on the basis of image texture features proposed in [6], in a two-step procedure. First, a set of texture features is extracted to represent an image block containing each crater candidate. Second, a simple and efficient classifier/feature selector is built using the AdaBoost learning algorithm to select features that better discriminate between craters and non-craters, and to assign probabilities of being a crater to each candidate. In this module users are asked to choose between different options related to the type of features (for instance, Haar-like or Histogram-of-Oriented-Gradients (HOG)), the dimension of the blocks to analyse or the number of weak classifiers to use. An example of a constructed dataset to be used in a training procedure is shown in Figure 3, where squares in green define positive examples (craters) and squares in blue mark the diversity of negative examples (non-craters).

5 Testing module

This module permits to map craters on a set of images using a model of classifier previously built. Users can select one single value for the probability threshold of a detected object being a crater or can define an interval and a step for evaluating different probabilities in a batch procedure. The evaluation of each classifier is done by the confrontation with ground-truth datasets. The typical performance rates (true and false detections, for positive and negative examples) and global derived indices are computed, as well as Receiver Operation Characteristic or ROC curves if desired. The outputs are delivered in diverse file formats permitting its direct use by other applications, namely those related to surface dating.

It is important to retain that this platform is intended to be mainly used for large scale practices, which means that after intensive training and testing phases, which should have led to the construction of robust classifiers, ground-truth datasets will be no longer available. That is, this module can be simply used for crater mapping.

6 Conclusions

The development of this application is still a work in progress but whose detection sequence, from the generation of candidates to the production of detection crater maps, is expected to be completed soon. The modular architecture of the platform will permit an easy inclusion of new approaches in every step of the processing sequence, for instance, to include additional type of filters, image features or classifiers.

References

Abstract

The baroreflex loop has a significant role in short-term blood pressure regulation. In order to keep cardiovascular homeostasis, the autonomic nervous system (ANS) controls the heart rate, arterial peripheral resistance and myocardium inotropy.

Several cardiovascular diseases and pathologies are associated to ANS dysfunctions, such as orthostatic hypotension, postural orthostatic syndrome and vasovagal syncope. The baroreceptor reflex and heart rate variability (HRV) are thus preferential targets for autonomic disorder tests. However, many parameters and relationships between different parts of the baroreflex system are still unknown.

In this paper, a physiologically based, mathematical model of the baroreflex loop is proposed. The predicted sympathetic and vagal dynamics, are compared with real data achieving a high degree of agreement. The real data was acquired from 7 healthy young subjects, who performed an active orthostatic test under continuous monitoring of the RR-interval.

The close fit between the predicted baroreflex pattern and the real physiological response, allows the estimation of reliable parameters and gains of sympathovagal balance.

1 Introduction

The control of cardiovascular system’s homeostasis is provided by the parasympathetic and sympathetic branches of the Autonomic Nervous System (ANS). Baroreflex is a negative feedback reaction of the human body to the environmental stimuli in order to keep stable arterial Blood Pressure (BP).

The baroreceptors, mechanical stretching receptors, located in the carotid sinus and aortic arch are responsible for the detection of BP oscillations. Changes in the activation pattern of these receptors are transmitted to the central nervous system areas, such as the nucleus of tractus solitarius (NTS), as shown in Fig 1. In the brainstem, the NTS, the nucleus ambiguus (NA) and the rostral ventrolateral medulla (RVLM) are the major control centers of parasympathetic and sympathetic activity [3, 4]. The NTS is the main integration center of stimuli from sensory receptors in the vasculature. This brainstem area has excitatory output to the RVLM, which increases sympathetic activity. Excitatory neuronal connection between NTS and NA also exist. The latter is responsible for the increase of parasympathetic (vagal) activity. In fact, NA gives root to branchial efferent motor fibers of vagus nerve activity [3, 4, 14].

Heart Rate (HR) is controlled by impulses originated in the Sinuatrial Node (SAN) - a fundamental part of the specialized excitatory and conductive system of the heart. The SAN myocytes owe their pacemaker activity to their capability of an automatic rhythmic discharge [5]. Acetylcholine (Ach) and Norepinephrine (NE) are released on the SAN and atrioventricular node (AVN) [4] by vagus and sympathetic nerves, respectively. NE increases the HR and forces myocardium contraction by the activation of beta1-adrenergic receptors. Furthermore, the increase of HR involves protein kinase A (PKA) and cAMP second messenger mechanisms [6, 10], raising the SAN myocytes excitability and the action potential rhythm. On the other hand, parasympathetic stimulation decreases the rhythm of sinus depolarization by Ach binding at the muscarinic receptor and rising of the conductance of the potassium channels [3, 4, 5, 9].

Heart Rate Variability (HRV) analysis is the preferred method for autonomic system assessment. Two main frequency bands [2] are usually considered in HRV spectral analysis: low frequencies (LF), from 0.04 to 0.15 Hz, and high frequencies (HF), from 0.15 to 0.4 Hz. The sympathetic activity can be observed in the LF band and the parasympathetic activity in the HF band. Changes in LF/HF ratio are related to sympathovagal balance.

In this work, an incremental baroreflex model for short-term HR and BP control is proposed. The model, is based on the previous work by Santos et al. [7]. The model parameters are re-estimated and a time delay, characteristic of the physiological process, introduced, see Table 1. A new physiological approach for the heart dynamics is also proposed.

2 Problem Formulation

In this work, an incremental baroreflex model for short-term HR and BP control is proposed. The model, is based on the previous work by Santos et al. [7]. The model parameters are re-estimated and a time delay, characteristic of the physiological process, introduced, see Table 1. A new physiological approach for the heart dynamics is also proposed.
pacemaker activity are modelled with a time delay and a low-pass filter, previously proposed in the *in vitro* study by Spear et al. [11]. Due to the different metabolic pathways of the parasympathetic and sympathetic systems, these stimuli suffer different time delays. Therefore, the release of Ach and NE in the SAN is affected by a time lag (Table 1), modelled by \( D_P \) and \( D_N \), respectively. The neurotransmitters’ dynamics are modelled by low-pass filters with coefficients \( \tau_P = 0.15 \) and \( \tau_N = 1 \) and unitary gain, following the physiological approach by [3, 11]. Parasympathetic activation increases the SAN depolarization period, rising the RR-interval (decrease of HR). On the other hand, sympathetic stimulation decreases RR-interval (increasing HR). Furthermore, in heart block, HR and sympathetic system stimuli are converted in the cardiac output (CO).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Variable</th>
<th>Value</th>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( D_P )</td>
<td>0.2sec</td>
<td>( G_P )</td>
<td>0.5</td>
<td>( D_N )</td>
<td>7.0sec</td>
</tr>
<tr>
<td>( D_P )</td>
<td>0.5sec</td>
<td>( G_P )</td>
<td>0.5</td>
<td>( G_P )</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Table 1: Model gain factors and time delay parameters used in simulation.

The CO and peripheral resistance \( (PR) \) are important parameters in Mean Arterial Pressure (MAP) control [3]. The Cardiovascular System (CVS) block only suffers the influence of the sympathetic system. The sympathetic gain in the CVS block is represented by \( G_Y \); this component is affected by a time delay, \( D_Y \), determined by the analysis of real data and simulation tests. This system can be characterized by an elastic component and a viscous component of the blood vessels’ walls [3, 7].

The baroreceptors block is composed by a low pass filter followed by a gain \( G_B \) (see Table 1). The physiological characteristics of the MAP signal in the aortic arch and carotid sinus imply the existence of a time delay, expressed as \( D_B \).

3 Experimental Results

An active orthostatic test (AOT) was simulated using the described model. The simulation was performed using a ramp form signal, mimicking the change of body position, corrupted with zero mean, low variance, Gaussian noise, representing the measurement’s noise. The MAP is the input signal, which is analysed by baroreceptor block.

The analysis of the model’s response provides a physiological based explanation for the HRV frequency bands discussed before. The physiological limits of the sympathetic and parasympathetic frequency are supported by the intrinsic properties of the SAN autonomic regulation, i.e., the dynamics of the neurotransmitters [11]. Hence, ANS modulation of the heart rate can be simulated by the response of LP filters, with the cut-off frequencies determined by Ach and NE kinetics. In this way, the cut-off frequency of the sympathetic system is 0.1730Hz and the parasympathetic cut-off frequency is 1.2631Hz, which falls within the generally accepted values [1, 2]. The simulation results, shown in Fig. 3, support the existence of a time lag, due to the sympathetic activation of smooth arterial muscles, included in the model in the CVS block.

In order to test this mathematical model, real RR-interval data was acquired during several AOTs. Seven healthy young subjects participated in this study, 4 women and 3 men. The AOT was chosen as the most exemplary evidence of baroreflex and ANS activation [2], instead of the passive orthostatic test (Head-Up Tilt Table Test) which cannot be easily analysed due to emotional reaction of subjects during the procedure. The experimental protocol was adapted from Cybulski et al. [15]. The test began with a 5 minute resting period in supine position. After the induction, the subject stands up quickly, in 2 – 4 seconds, and remains in the orthostatic position during 5 minutes. Active postural change from supine to the standing position causes a BP decrease and activation of negative feedback baroreflex are. The AOT response in subjects, represented in this work, was also similar to the described in the literature [2, 15]. Fourteen datasets of AOT were obtained, the average RR-interval variation is represented in Fig 3.

Some unknown parameters of baroreflex system were stimulated by empirical approach, e.g., the sympathetic gain and time delay in CVS for arterial constriction, such as the gain of baroreceptors. Another baroreflex characteristics were confirmed (i.e.,SAN dynamics) by comparing the average real data response with the simulation’s results. Fig 3 shows the real and simulated RR-interval variation during the AOT, the results are in agreement with the results predicted by Spear et al. [11]. The main difference between the real and predicted curves lies in the recovery time after the maximum sympathetic activation. This difference is expected since the baroreceptor response is not linear with decreasing and rising BP. The cardiorespiratory coupling was not considered [13, 14], which can introduce variability in baroreflex response. Nevertheless, the level of agreement between the results suggests that human short-term blood pressure dynamics can be approximated by a baroreflex incremental model.

4 Conclusions

The Baroreflex system is an essential short-term control mechanism of BP, associated with many autonomic diseases and pathologies. Understanding the physiological processes of the baroreceptor reflex allows the design of realistic mathematical models, useful to compute accurate parameters and gains and predict the system’s behaviour under new conditions.

The incremental baroreflex model proposed in this paper follows the behaviour expected for the human system. The performed simulations provide estimations for previously unknown parameters, such as the neurotransmitters dynamics in SAN.

The obtained results show an agreement between the simulations and real data. While the response time is virtually identical, the recovery dynamics are slightly different, suggesting that a non-linear baroreceptor model might yield better results. The differences resides in the recovery dynamics, possibly due to individual physiological differences between subject and respiratory influence.

In conclusion, the baroreflex model is a useful tool to study short-term cardiovascular dynamics and ANS activity. Further work will try to introduce cardiorespiratory coupling in the baroreflex model in order to mimic the sinus arrhythmia.

References


3D Texture Analysis using Local Binary Patterns

Pedro M. Morgado
pedromorgado@isr.ist.utl.pt

Margarida Silveira
msilveira@isr.ist.utl.pt

Jorge S. Marques
jsm@isr.ist.utl.pt

Instituto Superior Técnico
Institute for Systems and Robotics
Lisbon, Portugal

Abstract

Texture analysis is a tool commonly used for object recognition in 2D images. However, 3D texture descriptors have been far less investigated. In this paper, we evaluated a previously proposed generalization to 3D of a texture descriptor known as Local Binary Patterns (LBPs), conducting a comparative study in two distinct problems and using three other approaches: standard LBPs applied to slices of the 3D image; histograms of the 3D image’s gradient; and 3D Haar-like features. Results obtained on synthetic data showed that all descriptors can easily encode textural information. On real data, the proposed method was able to achieve performances similar to Haar features, but superior to the other two.

1 Introduction

Texture plays an important role in the human recognition of different objects. As for computer vision applications, even though there is no single accepted mathematical definition to characterize it, the extraction of texture features has already revealed to be very useful. Recently, with the emergence of new technologies that allow the acquisition of 3D volumetric data, several texture extraction algorithms have been extended.

This paper is mainly focused on the descriptor known as Local Binary Patterns [8], for which we previously introduced a novel generalization that successfully extended LBPs to volumetric data, including both uniformity and rotation invariance concepts [7]. More concretely, we will compare this method with three other approaches: the first is simply the 2D LBPs as proposed in [8] and applied to slices of the volume; the second is a generalization of an alternative texture descriptor that uses histograms of the gradient of the image as features; and the third is based on 3D Haar-like features. The implementation of the last two procedures followed closely Bicacro’s work [1]. Moreover, two datasets (one synthetic and one real) were used to perform the comparison.

2 Methods

2.1 2D Local Binary Patterns

LBPs were originally proposed by T. Ojala et al. [8] to perform texture analysis on 2D images. According to this work, textures can be identified by the relative frequency of occurrence of certain patterns called LBPs, which are defined as follows. Consider first, for a given pixel $x_k$, a set of $P$ neighbors, $\{x_1, \ldots, x_P\}$, located over a circumference of radius $R$. The LBP associated with pixel $x_k$ can be extracted by comparing the gray-levels of the neighbors, $\{V_1, \ldots, V_P\}$, with its own gray-level, $V_c$. The outputs of such comparisons (i.e. 1 if $V_p > V_c$ or 0 otherwise) are joined into a binary vector $T$ that uniquely represents the extracted pattern.

Although there is a total of $2^P$ possible patterns, most of them occur very rarely on an image, and thus the estimates of their probabilities of occurrence are not robust. To avoid this problem, a class of frequent patterns called uniform LBPs was introduced and all non-uniform LBPs were merged into the same group. According to [8], uniform patterns should have no more than two transitions from 0 to 1 or vice-versa when traversed circularly. The concept of rotation invariance was also introduced in [8], which allowed for the identification of the same texture with different orientations. It was accomplished by joining into the same group all patterns that can be aligned after some appropriate rotation.

2.2 3D Local Binary Patterns

In this section, we will describe briefly our extension of LBPs to 3D data. The interested reader should report to [7] for an extended and more detailed description of the proposed generalization, including implementation details for faster texture analysis and a more comprehensive study of the parameters involved. The generalization of the basic LBP concept is straightforward. We only need to define for each voxel a neighbor set, where the neighbors now lie evenly distributed over a sphere, instead of a circumference. In the absence of known regular polyhedra with the desired number of vertices, numerical solutions, such as the ones proposed in [4], can be used to define the neighbors coordinates.

The generalizations of the uniformity and rotation invariance concepts constitute a more challenging problem.

Uniformity Since the original definition of uniformity cannot be easily extended to 3D, we proposed a new but equivalent definition. Consider first two sets of neighbors: one formed by the neighbors $x_p$ with gray-levels greater than the gray-level of the central voxel ($V_p > V_c$), and the other formed by the remaining ones. According to the proposed definition, an LBP is said to be uniform if the convex hulls of these two sets of points do not intersect. Note that this definition can be directly applied to any dimension, and when used in the original 2D setting, the same notion of uniformity is achieved, i.e. the same patterns are labeled as uniform.

Rotation Invariance A look-up table that maps each pattern to a rotation invariant LBP label was created as follows. First, each LBP is engraved into a spherical function $f(\theta, \phi)$ that assumes the value 1 at a small neighborhood of every point of the pattern for which $V_p > V_c$ and 0 everywhere else. Then, this function is decomposed into a sum of spherical harmonics, using a maximum degree of expansion $l_{max}$ as in

$$f(\theta, \phi) \approx \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} a_{lm} Y_l^m(\theta, \phi),$$

where $Y_l^m$ is the spherical harmonic base function of degree $l$ and order $m$ and $a_{lm}$ the respective complex coefficient. Groups of rotation invariant LBPs can then be identified by the following rotation invariant shape descriptor, which was proposed in [5]:

$$SH = \{ \| \pi_l(\theta, \phi) \|, \ldots, \| \pi_{l_{max}}(\theta, \phi) \| \},$$

where $\pi_l = \sum_{m=-l}^{l} a_{lm} Y_l^m$ is the projection of the function $f$ onto the subspace formed by the span of spherical harmonics base functions restricted to a given degree $l$. In order to allow small errors, two LBPs were considered to belong to the same (rotation invariant) group if the distance between the corresponding vectors $SH$ were below a predefined value $\eta$.

2.3 Histograms of Gradient Magnitude and Orientation

The gradient of an image can be an important source of information for texture discrimination. Following the implementation of Bicacro et al. [1], the three spatial derivatives of the image $V(\partial V/\partial x, \partial V/\partial y, \partial V/\partial z)$ were first approximated using symmetric first-order discrete differences. Then, three gradient quantities (magnitude $\| \cdot \|$, azimuth $\phi$ and polar $\theta$ angles) were computed for each voxel using Eq. 3 to 5 and histograms of these quantities were computed and used as features.

$$\| \nabla V \| = \sqrt{(\partial V/\partial x)^2 + (\partial V/\partial y)^2 + (\partial V/\partial z)^2}$$

$$\phi = \arctan \left( \frac{\partial V/\partial y}{\partial V/\partial x} \right)$$

$$\theta = \arctan \left( \frac{\sqrt{(\partial V/\partial z)^2 + (\partial V/\partial y)^2}}{\partial V/\partial x} \right)$$
2.4 3D Haar-like Features

The formulation of the Haar-like features used in this work was proposed by Cui et al. [3]. It relies on simple structures such as the ones represented in Fig. 1. More specifically, features are computed by placing these structures at predefined positions of the original volume, summing the voxels inside the lighter regions and subtracting those inside the darker ones. Moreover, different dimensions of the structures are typically allowed, so that patterns can be searched at different scales.

Figure 1: Types of 3D Haar-like structures used to extract features.

2.5 Feature Selection and Classifier

After the feature extraction stage, the dimensionality of the feature vector was reduced using a ranking selection algorithm based on the Mutual Information between each feature and the class label. Then, the classification stage was conducted with a multi-class implementation of Support Vector Machines (SVM) [2], using a “one-against-one” approach. In this paper, only results with a linear kernel are presented but other kernels were also tested, achieving worse or comparable performances.

3 Experiments

3.1 Datasets

The four approaches were compared on real and synthetic data.

An artificially generated database composed by 2048 samples evenly distributed over four classes of solid textures with dimensions of 64 × 64 was constructed using Perlin noise [9]. In order to harden the classification problem, we added zero mean Gaussian noise to all generated samples, and also increased intra-class variability by allowing a small randomness in the parameters that define each class.

To test our method on real data, we used Positron-Emission Tomography images from the Alzheimer Disease Neuroimaging Initiative (ADNI) database. The images used belonged to three classes – Alzheimer’s Disease (AD), Mild Cognitive Impairment (MCI) and Normal Controls (NC) – and had already undergone a series of pre-processing steps to reduce meaningless differences between different patients as described in [6].

3.2 Results

Tables 1 and 2 present several confusion matrices (one for each type of feature) as well as the overall accuracies. The true classifications are associated with rows of these matrices and predictions with their columns.

The synthetic database was easily classified by the three approaches (Table 1). The proposed extension of LBPs to 3D achieved the best results, misclassifying only 6 samples out of 2048, 4 samples less than its 2D counterpart, and 5 samples less than the HGMO approach. In fact, after the classification process, we analyzed the samples misclassified by the three approaches and the results were found in the overall performances.

The results associated with the automated diagnosis of AD and MCI are presented in Table 2. As it can be seen, the MCI state is the hardest to classify since it is an intermediate state between NC and AD. Nevertheless, all extraction techniques were able to perform better than a random classifier within this class (all rates of correct predictions of MCI patients are higher than 33%). Regarding the generalization of LBPs to 3D, this novel approach outperformed significantly its 2D counterpart and the HGMO extraction procedure, and obtained an overall accuracy identical to 3D Haar-like features. The good results were attained especially due to its higher rate of correct classifications of normal controls, although being outperformed in the classification of MCI and AD by 2D LBPs and Cloud-LBPs, respectively.

Table 1: Confusion matrices and overall accuracies attained in the classification of the synthetic database. All measures are given in [%].

<table>
<thead>
<tr>
<th></th>
<th>Cloud</th>
<th>Fire</th>
<th>Marble</th>
<th>Wood</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D LBPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud</td>
<td>99.6</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>99.7</td>
</tr>
<tr>
<td>Fire</td>
<td>0.2</td>
<td>99.8</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Marble</td>
<td>0.6</td>
<td>0</td>
<td>99.4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wood</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2D LBPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud</td>
<td>99.8</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>99.5</td>
</tr>
<tr>
<td>Fire</td>
<td>0.4</td>
<td>99.6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Marble</td>
<td>0.8</td>
<td>0.6</td>
<td>98.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wood</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>HGMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud</td>
<td>99.0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>99.5</td>
</tr>
<tr>
<td>Fire</td>
<td>0.4</td>
<td>99.6</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Marble</td>
<td>0.8</td>
<td>0</td>
<td>99.2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wood</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Confusion matrices and overall accuracies attained in the classification of the ADNI dataset. All measures are given in [%].

<table>
<thead>
<tr>
<th></th>
<th>NC</th>
<th>MCI</th>
<th>AD</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D LBPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>79.7</td>
<td>16.9</td>
<td>3.4</td>
<td>65.0</td>
</tr>
<tr>
<td>MCI</td>
<td>25.4</td>
<td>50.9</td>
<td>23.7</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>3.4</td>
<td>32.2</td>
<td>64.4</td>
<td></td>
</tr>
<tr>
<td>2D LBPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>62.7</td>
<td>33.9</td>
<td>3.4</td>
<td>58.8</td>
</tr>
<tr>
<td>MCI</td>
<td>28.8</td>
<td>44.1</td>
<td>27.1</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>6.8</td>
<td>23.7</td>
<td>69.5</td>
<td></td>
</tr>
<tr>
<td>HGMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>69.5</td>
<td>25.4</td>
<td>5.1</td>
<td>61.6</td>
</tr>
<tr>
<td>MCI</td>
<td>25.4</td>
<td>50.8</td>
<td>23.8</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>16.7</td>
<td>34.0</td>
<td>52.5</td>
<td></td>
</tr>
<tr>
<td>Haar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>74.6</td>
<td>22.0</td>
<td>3.4</td>
<td>65.0</td>
</tr>
<tr>
<td>MCI</td>
<td>23.8</td>
<td>54.2</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td>1.7</td>
<td>32.2</td>
<td>66.1</td>
<td></td>
</tr>
</tbody>
</table>

4 Conclusion

This paper evaluated a recently proposed extension of LBPs to 3D. The new generalization is able to replicate closely the three most important concepts of the original 2D descriptor: the construction of LBPs, uniformity and rotation invariance. The proposed method was also compared with three other texture descriptors and tested on real and synthetic data. The experimental results revealed that the proposed 3D extension was able to outperform the standard LBP features when extracted from slices of the volumetric image, as well as the HGMO type of feature. As for the Haar-like features, similar performances were attained.

Acknowledgments

This work was supported by Fundação para a Ciência e Tecnologia through the ADIAR project (PTDC/SAU-ENB/114606/2009) and the project PEst-OE/EEI/LA0009/2013.

References

Automatic sleep parameter computation from Activity and Cardiovascular data

Alexandre Domingues1
adomingues@gmail.com

Teresa Paiva2
teresapaiva0@gmail.com

J. Miguel Sanches1
jmrs@ist.utl.pt

1Instituto de Sistemas e Robótica
Instituto Superior Técnico
2Centro de Electroencefalografia e Neurologia Clínica
Faculdade de Medicina da Universidade de Lisboa

Abstract

Automatic computation of sleep parameters from data acquired with portable sensors, is a challenging problem with important clinical applications. In this paper, the Sleep Efficiency, and REM and Non-REM sleep percentages are automatically computed from ECG, Respiration and Actigraphy. The algorithm relies on two classifiers, designed to reject ambiguous data, and a regularization step that corrects the predicted number of samples on each of the considered classes. The described method achieves an estimation error of 4.5%, 9.9% and 5.5% on Sleep Efficiency, REM and NREM percentages respectively.

1 Introduction

Automatic computation of sleep parameters (SP), from data acquired in mobile environments, is an open issue that poses many challenges. Several approaches have been proposed to properly identify the different sleep stages, spectral analysis of the HRV plays a major role in many publications, where the frequency bands described in [3] have become the standard for HRV spectrum analysis. The estimation of sleep parameters/stages from multi-modal data is presented in some papers with promising results. In [4] the authors combine ACT and Cardio-Respiratory signals achieving high accuracies in Sleep and Wakefulness detection, although no proper validation data, (i.e. the hypnogram from the PSG), is used. In [5] the authors combine HRV with respiratory inductance plethysmography (RIP), achieving a k-index [1] of k = 0.45 for a 3 class discrimination (Wake, REM, NREM) and k = 0.57 for a simplified 2 class problem (Wake,Sleep).

This paper deals with the problem of automatically estimating three standard SP: i) SE, ii) REMp and iii) NREMp, from ECG, Actigraphy (ACT) and RIP, which are easily acquired with portable sensors. The described method eliminates the need of a full Hypnogram by combining the rejection of ambiguous samples and a regularization operation.

2 Methods

The SP estimation method, displayed in Fig. 1, is composed by the pre-processing and feature extraction procedures, followed by a classification stage, designed to reject ambiguous features, and a regularization operation. The multi-modal dataset used in this work includes ECG and RIP data, obtained from PSG and ACT, acquired with a wrist actigraph, from 15 healthy volunteers.

The pre-processing operations are required to reduce the movement artefacts, normalize the data across different patients and prepare it for feature extraction. After QRS complex detection, the RR signal is constructed from the detected R peaks, and downsampled to 2 Hz. Magnitude normalization and DC component removal are applied to both RIP and ACT signals.

After pre-processing, each dataset is divided in contiguous epochs of T = 30 seconds, synchronized with the ground-truth hypnogram provided by the medical staff. All the epochs corresponding to any of the 3 distinct Non-REM sleep stages were grouped into one single label.

The feature vector includes the standard frequency and time domain RR features described in [3], the ACT features from [2] and mean and standard deviation of the respiratory rate. The statistical significance of the included features was assessed with a one-way ANOVA test.

The classification step relies on two classifiers that independently classify all samples into i) Sleep/Wakefulness (SW) and ii) REM/Non-REM. Each classifier is designed to take into account a rejection factor (RF), rejecting a specified percentage of samples, whose classification is ambiguous. The rejection works by computing the true or estimate posterior probability of the winning class for each sample and rejecting those which are below the specified percentage.

Each classifier maps every sample into one of three labels: SW ∈ {sl, wk, r} and RN ∈ {rs, ns, r} where rs, ns, wk and r refer to REM, Non-REM, Wakefulness, Sleep and Rejected sample respectively.

During the training step, the two binary classifiers are trained with data from the 2 considered classes. However, during the test, they map samples belonging to 3 classes. Any sample from a class not predicted by the classifier will either be miss-classified or rejected. The Support Vector Classifier (SVC) with a quadratic kernel was chosen for both classification tasks.

Let us consider a binary classifier C, with a reject option, which maps each sample into one of three labels l ∈ {p, n, r} where p, n and r denote Positive, Negative and Reject.

The confusion matrix1 is represented as

\[
A = \begin{bmatrix}
Tp & Fn & Rp \\
Fp & Tn & Rn
\end{bmatrix}
\]

where Tp, Fn, Fp, Tn and Rn are the True Positives, False Negatives, False Positives, True Negatives, Rejected Positives and Rejected Negatives respectively.

The positive (θp,i) and negative (θn,i) correction factors and the fraction of rejected samples per class (wp,i) and (wn,i) are computed for each

\[1\text{The positive detection rate is computed as } \frac{Tp}{Tp + Fn} \text{ and the global accuracy is } \frac{Tp + Tn + Rn}{Tp + Fn + Fp + Tn + Rn}\]
training dataset as
\[
\theta_{p,i} = \frac{\text{TP}_p + \text{FP}_p}{\text{TP}_p + \text{FN}_p}
\]
\[
\theta_{n,i} = \frac{\text{FN}_n + \text{TN}_n}{\text{FP}_n + \text{TN}_n}
\]
\[
\omega_{p,i} = \frac{\text{RP}_p}{\text{RP}_p + \text{RN}_p}
\]
\[
\omega_{n,i} = \frac{\text{RN}_n}{\text{RP}_p + \text{RN}_n}
\]
with \(i \in \{1, \ldots, M\}\), and \(M\) the number of training datasets. The final values are obtained averaging over \(\theta_{(p,n)}\) and \(\omega_{(p,n)}\).

The number of estimated samples on each class can be improved by correcting the number of predicted samples as
\[
N(p) = \frac{N(p)}{\theta_p}
\]
\[
N(n) = \frac{N(n)}{\theta_n}
\]
where \(N(\cdot)\) is a counting operator, and estimating the number of rejected samples from each class as
\[
N(r_p) = \omega_p N(r)
\]
\[
N(r_n) = \omega_n N(r)
\]

The expressions for the three SP can now be written as
\[
\text{SE} = \frac{N(s) - \theta_p N(s) + \omega_p N(wk))}{\theta_p}
\]
\[
\text{NREM}_p = \frac{N(ns) + \omega_p N(rns)}{N(ns) + N(rs) + N(r) + \text{SE}}
\]
\[
\text{REM}_p = \frac{N(ns) + N(rs) + N(r) + \text{SE}}{N(ns) + N(rs) + \text{SE}}
\]
where \(SE\) is computed from the output of the SW classifier and \(\text{NREM}_p\) and \(\text{REM}_p\) from the RN classifier.

## 3 Results

Each subject performed a standard nocturnal PSG exam at a sleep laboratory. The PSG data was jointly acquired with ACT using a Somn-o-watch\textsuperscript{TM} device, from Somnomedics, placed in the non-dominant wrist of the subjects, acquiring with a sampling rate of 1Hz. The core of these devices is a 3D accelerometer that measures the acceleration along 3 orthogonal axis with a configurable output format. Here, the output of the actigraph is the acceleration magnitude.

The hypnogram, obtained from the PSG by trained technicians, is used as a ground truth to identify REM sleep, Non-REM sleep and wakefulness in epochs of 30 seconds. Fifteen adult subjects (age 44.4 ± 11 years, 10 Males, 5 Females), with no pre-diagnosed sleep disorders, participated in this study.

The SE was computed from the hypnogram for each patient. All the values of SE fell within the range 85% – 95%. These values are within the accepted range for healthy subjects, usually above 85% [6].

The three SP and the estimation error\textsuperscript{2} were computed, for each dataset using the proposed method using a leave-one-patient-out cross-validation, where each patient dataset is tested after training the algorithm with the remaining data. Table 1 shows the average value and error for each parameter, computed for several different RFs. Using a RF of 10% the average values are almost coincident with the real values. The estimation errors are 4.5% for the SE, 9.9% for the REM\(_p\) and 5.5% for the NREM\(_p\).

In order to test the influence of the training and test sets and to assess the generalization capability of the algorithm the following steps were performed:

1. Ten datasets were randomly selected from the pool of 15 available datasets.
2. From these 10 datasets, 5 were randomly selected to train the algorithm.
3. The SP were estimated for the remaining 5 datasets and the average error computed.

This procedure was repeated 10 times resulting in average errors of 5.7 ± 1.2, 11.9 ± 4.3 and 4.4 ± 2.3 for SE, REM\(_p\) and NREM\(_p\) respectively. These values are very similar to the ones reported in Table 1 suggesting that the reported results should be extensible to other datasets.

## 4 Conclusion

In this paper we propose a new method to estimate sleep parameters from RR, RIP and ACT data. The method relies on two classifiers designed to reject ambiguous data and a regularization parameter based on the a priori informations from the classifiers performance and rejection patterns.

With this method the estimation errors are \(\approx 5\%\) for SE and NREM\(_p\) and \(\approx 10\%\) for REM\(_p\).

## Acknowledgement

This work was supported by the FCT project [ PTESOE/EIE/LA0009/2013] and FCT project "Detection of Brain Micro-states in Fibromyalgia (PTDC/SABEB/104948/2008)"

## References


Supervised Feature Discretization with a Dynamic Bit-Allocation Strategy

Artur J. Ferreira
artuj@isel.pt

Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt

Instituto de Telecomunicações, and
Instituto Superior de Engenharia de Lisboa
Lisboa, PORTUGAL
Instituto de Telecomunicações, and
Instituto Superior Técnico
Lisboa, PORTUGAL

Abstract
The use of feature discretization (FD) is useful in several machine learning and pattern recognition tasks. By attaining adequate and compact data representations, FD may improve the performance of many methods. It is often the case that learning with discrete data representations yields both lower training time and better accuracy, as compared to the use of the original features. Moreover, FD may also allow for a better human understanding/interpretation of the data. However, many FD techniques are sub-optimal, in the sense that they do not take into account feature interdependencies, as discretization is carried out.

In this paper, we propose a dynamic supervised FD technique, addressing feature interdependencies. Our method selects the discretization cut-points by simultaneously maximizing two criteria: the dependency between the discretized features and the class label; the independence among these features. The method performs an incremental bit allocation scheme, using mutual information (MI) as the dependency/independence measure. Experimental results on low and medium-dimensional datasets show that the proposed method often achieves better accuracy than other well-known supervised FD approaches.

1 Introduction
In machine learning and pattern recognition problems, it is common to find datasets that contain both categorical and numeric features. Whereas the categorical features are already discrete, the numeric features require integer or floating point representations. For these numeric features, some machine learning and pattern recognition tasks can be made easier by performing feature discretization (FD). FD seeks compact data representations, desirable ignoring minor fluctuations that are irrelevant for the task at hand, leading to more robust classifiers. In some cases, FD is even mandatory in order to be able to use some learning procedure, or it is intrinsically built into the structure of the classifier and the corresponding learning algorithm (e.g., in classification trees).

The literature on FD is vast, with a wide range of unsupervised and supervised techniques; comprehensive reviews can be found in [3, 6, 8, 10]. The recent survey [6], identifies more than 80 FD methods, most of which are supervised (i.e., they use the class labels in the training data).

In this paper, we propose a dynamic supervised FD technique, which performs FD taking into account the feature interdependencies. For a given feature, the discretization cut-points are computed as functions of the dependency with both the class label and the already discretized features. The method relies on the use of the mutual information (MI) [2] criterion to assess these (in)dependencies.

The remainder of this paper is organized as follows. Section 2 briefly reviews existing supervised FD techniques. Section 3 presents our dynamic supervised method for FD. The experimental evaluation is reported in Section 4. Finally, Section 5 presents some concluding remarks.

2 Related Work on Feature Discretization
This Section briefly reviews some successful supervised FD techniques, based on information-theoretic and statistical criteria. In this context, we also describe our previous approach for supervised FD, based on a bit allocation scheme.

2.1 Well-known Approaches
The information entropy minimization (IEM) [4] and its variant IEM variant (IEMV) [7] are successful supervised FD methods; these approaches use information-theoretic criteria, namely the entropy minimization heuristic, for discretizing continuous values into multiple intervals, leading to small decision trees to classify the data. The class-attribute interdependence maximization (CAIM) [9] algorithm uses a statistical approach, aiming to maximize the class-attribute interdependence and to minimize the number of discretization intervals. A related method is the class-attribute contingency coefficient (CACC) [11], which is based on the maximization of a modified version of the contingency coefficient. This modification overcomes the key drawbacks of earlier schemes such as CAIM. The main contribution of CACC is that it can generate a better discretization scheme leading to the improvement of decision-tree classifiers.

2.2 Mutual Information Discretization Pool (MIDP)
Recently, we have proposed a bit-allocation strategy for supervised FD based on the maximization of the MI between each discretized feature and the class label [5]. The proposed mutual information discretization pool (MIDP) method starts with a pool of \( P \) bits to be distributed among the set of \( d \) features, and then proceeds as follows:

1) each feature \( X_i \) has a counter \( b_i \) holding the number of bits currently allocated to it; these counters are initialized at zero;

2) for each feature, perform incremental discretization with \( b_i + 1 \) bits, maximizing an estimate of the MI between the resulting discrete feature \( \bar{X}_i \) and the class label \( Y \): \( MI(\bar{X}_i; Y) \);

3) at each iteration, assign one more bit \( (b_i \leftarrow b_i + 1) \) to the feature that exhibits higher increase in the MI \( MI(\bar{X}_i; Y) \);

4) the previous step is repeated until either \( a = \sum_{i=1}^{d} b_i \) reaches \( P \) (no more bits in the pool) or no feature exhibits an increase in the MI.

Figure 1 depicts an example of the use of the MIDP technique, for \( d = 5 \) features and \( P = 8 \) bits. For each feature, we display the values of MI using \( q \in \{1, 2, 3, 4\} \) bits. The number in (blue) italic font at the right-hand-side of the MI value is the \( a \) counter mentioned above. At the bottom of the figure, we have the final number of bits \( b_i \) allocated per feature. Since MIDP does not take into account feature interdependencies, it is categorized as a static FD method, being able to perform well on high-dimensional data [5].

3 Dynamic Mutual Information Discretization
In this Section, we describe a new proposal for supervised FD: mutual information discretization pool conditional (MIDPC). In order to account for feature interdependencies, which often happen in many datasets, we propose a dynamic FD method. The method is also based on a pool of bits, as discussed above for the static MIDP algorithm.

![Figure 1: MIDP for \( d = 5 \) features and \( P = 8 \) bits. The number at the right-hand-side of the MI values is the sequence number of the allocated bits, \( a \), whereas \( b_i \) is the final total number of bits allocated per feature.](image-url)
Table 1: The datasets used in the experiments; \( d \), \( c \), and \( n \) are the number of features, classes, and instances, respectively.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>( d )</th>
<th>( c )</th>
<th>( n )</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Iris</td>
<td>4</td>
<td>3</td>
<td>150</td>
<td>Type of iris plant</td>
</tr>
<tr>
<td>#2 Balance</td>
<td>4</td>
<td>6</td>
<td>265</td>
<td>Psychological results</td>
</tr>
<tr>
<td>#3 Car</td>
<td>6</td>
<td>4</td>
<td>1728</td>
<td>Car acceptability</td>
</tr>
<tr>
<td>#4 Bupa</td>
<td>6</td>
<td>2</td>
<td>345</td>
<td></td>
</tr>
<tr>
<td>#5 Yeast</td>
<td>8</td>
<td>2</td>
<td>1484</td>
<td>Localization site of protein</td>
</tr>
<tr>
<td>#6 Wine</td>
<td>13</td>
<td>3</td>
<td>178</td>
<td>Classify wine cultivar</td>
</tr>
<tr>
<td>#7 Hepatitis</td>
<td>19</td>
<td>2</td>
<td>135</td>
<td>Hepatitis detection</td>
</tr>
<tr>
<td>#8 Dermatology</td>
<td>34</td>
<td>6</td>
<td>538</td>
<td>Skin disease diagnosis</td>
</tr>
</tbody>
</table>

MIDPC proceeds in a similar fashion as MIDP, but with a key difference: it selects the discretization cut-points that maximize the MI of the discrete features (the relevance) and minimize the dependency (the redundancy) among the discrete features. It performs a conditional discretization mechanism, since it decides that the feature that will receive the next bit in the pool, depends on all the previously discretized features. In summary, MIDPC proceeds with the same steps as MIDP, but with the key difference that it scans all the features, assigning one more bit to the feature that maximizes a rank function

\[
r_j = MI(\tilde{X}_i; Y) - \frac{1}{n_a} \sum_{k=1}^{n_a} MI(\tilde{X}_i; \tilde{X}_k),
\]

(1)

where \( n_a \) is the number of features for which at least one bit is allocated (i.e., for which \( b_i \neq 0 \)). Other possible rank criteria are

\[
r_i' = MI(\tilde{X}_i; Y) \left( \eta + \frac{1}{n_a} \sum_{k=1}^{n_a} MI(\tilde{X}_i; \tilde{X}_k) \right)^{-1},
\]

(2)

with \( \eta > 0 \) being a small positive constant, and

\[
r''_i = MI(\tilde{X}_i; Y) + \frac{1}{2\eta} \left( \sum_{k=1}^{n_i} H(\tilde{X}_i|\tilde{X}_k) + \sum_{k=1}^{n_i} H(\tilde{X}_k|\tilde{X}_i) \right),
\]

(3)

with \( H(\cdot|\cdot) \) denoting the conditional entropy [2]. The rationale behind these rank measures is that we aim at maximizing the MI between the discrete feature \( X_i \) and the class label \( Y \), while, simultaneously, minimizing the dependence between the discrete feature and all the previously discretized features.

Naturally, the MIDPC method has a higher computational cost than MIDP. In order to minimize this increase on the computational cost, for each feature we start by sorting its values and we remove existing duplicate values. Then, we perform a sequential search using each unique value as a candidate cut-point.

5 Conclusions

In this paper, we have proposed a dynamic supervised FD technique based on a combination of two criteria: the maximization of the mutual information between the discretized features and the class label and the minimization of the dependence among the discretized features. The classifiers learned on the features discretized by our method usually attain better accuracy than those learned on the original features or on the features discretized by other methods. The main drawback of our method is its high computational cost, which scales quadratically with the number of features, making it too costly for most high-dimensional datasets.

As future work, we will develop methods that scale more favorably for higher-dimensional datasets, namely by combining discretization with feature selection techniques and improving the cut-point search strategy. We will also explore modifications of our ranking criteria, namely by considering alternative measures of mutual information and entropy (e.g., Renyi’s and Tsallis’).

References

Abstract
Breast cancer is the most common cancer among females. Two main approaches are used as treatment: mastectomy, in which the cancerous breast is completely removed; and conservative treatment, in which the tumour is removed with margin of healthy tissue. To improve surgical approaches resulting less damage to dynamic shape of breast, it is worth to study the breast model to enable specialists having full comparison between the results of different treatments. The aim of this work is to study 3D reconstruction based on passive and active sensors. Also, it is aimed to study the state of the art about parametric models to obtain breast shape. Such parametric model can enhance surgeons experience in order to perform better surgeries and patients to be more confident about the breast shape after treatment.

1 Introduction
Nowadays, breast cancer is the most widespread cancer among women accounting for near 23% of all cancers and nearly 13.8% of cancer deaths in women; however, this issue is curable if it is detected in early stages. Based on diagnosing time, there are various treatments to remove tumours. [1].

Surgical methods determine the amount of breast tissue that should be removed, which to finally result in distorted breast shapes [2]. In Lumpectomy (which is known Breast Cancer Conservative Treatment or BCCT), just the tumour and a thin layer of healthy tissue around it are to be removed. It is essential for the patients to undergo a period of radiotherapy to ensure that all cancerous cells are removed perfectly [1]. However, in mastectomy total tissues of the breast are evacuated, which is resulted in huge loss of body. Moreover, the success of both methods is near equal [1].

Based on treatment, the breast is distorted due to tissue removal. Almost all patients feel unpleasant after surgery since the breasts are symbol of feminine. Obtaining complete and parametric 3D model of the breast has the advantage of breast shape prediction. Following a complete reconstruction, both surgeons and patients can intercept the changes that are caused by surgery. The result would enable the prediction of the shape of the breast. In this paper, we aim to review state of the art about breast 3D reconstruction and further, the parameterization of 3D models.

2 Human Body Reconstruction
Reconstructing 3D model of human body has enhanced measurement, evaluation and planning that are required for treatments. Unfortunately common human body reconstruction methods lack accuracy and are incomplete. Moreover they require expensive device which are not available in all clinics. For instance, proposed method in [3] provides accurate body shape reconstruction, however it requires laser scanners that costs a lot. In the other research reported in [4], it is proposed to perform body reconstruction with low-cost IR sensors; but their model had lacked in accuracy in medical diagnosis applications.

In 2011, Cui et al. [5] proposed a method to perform the reconstruction of human body by using one Kinect. Since they were not able to handle non-rigid movements, the reconstructed limb model such as arms and legs were poor in quality. In another work done by them recently in [6], they proposed a method that indicates they have coped with the problem by using global rigid and non-rigid alignments. They also improved the depth data quality by applying super resolution techniques on the model. At the same time, Weiss et al. [7] reconstructed human body by fitting the parameters of SCAPe in [8] that predefined model to depth data and image silhouettes.

3 Breast 3D Reconstruction
The idea of using 3D breast models in BCCT aesthetical evaluation motivated researchers to propose methods for breast reconstruction. Figure 1- Blaniuk et al. virtual reality approach [12]

Figure 1- Blaniuk et al. virtual reality approach [12]

Catanuto et al. [9] carried out a research on a set of parameters by which they determined the shape of breast; for both healthy and treated breast using an optoelectronic tracking system. They reconstructed patient’s perspective self-view to present divergence angle. On the other hand they used colorful map indicating flat regions or curvatures for introducing other features. They also used a real-time breathing artifact correction automatically to make the method more robust.

Eder et al. [10] studied the 3D evaluation protocols to analyze asymmetry via the differences between both breasts. They used a 3D scanner to reconstruct the breast and finally showed that the current method could assist surgeons both in pre-operative planning and breast reconstruction optimization after plastic surgeries.

Henseler et al. [11] proposed the usage of multiple stereo camera system to reconstruct breast model without human interaction. Using a stereo photogrammetry consisting of 8 cameras located on 4 pods two by two, they constructed breast models from dummy torso in both Figure 2-a and Figure 2-b.

4 Breast 3D Parametric Model
Sole reconstruction could be used in order to perform measurement, but using a complete parametric model, all measurement, prediction and planning would be available. Highlighting the necessity of a tool for both surgeons and patients to predict results of the surgery, Balaniuk et al. [12] combined both virtual reality approaches and soft tissue modeling methods to simulate reconstruction or augmentation surgeries using 3D tools. AS shown in Figure 1, both distorted and implanted breasts are modeled. Kovacs et al. [13] also introduced their work on to measuring breast volume by 3D scanner. They studied scanner-related factors and tested the most favorable imaging technologies on dummy models to guarantee that the method is reproducible.

In another research to reconstruction of human body, predefined models (known as avatar) are implemented that equipped with scaling parameters. In Figure 3, three avatars are considered as initial model; male, female and child. Determining the initial model, the reconstruction continues to change scaling parameters to fit the captured data on a correct avatar [14]. Also the proposed method is not automatic in detection of the initial avatar.

In none of the above mentioned methods, it was proposed a complete 3D parametric model. Besides, they are not either accurate enough or require expensive devices to be used in clinics.
5 Conclusion

Common 3D reconstruction methods have some difficulties in implementation due to high cost devices, requiring expert personnel and computation complexity; hence they have not proposed a complete breast parametric model. Having a complete parametric model, surgeons will be able to plan for required treatments. Moreover the breast shape would be predicted before performing any surgery. Based on planning and prediction facilities, specialists are able to improve both the effectiveness and the cosmetic result of the surgeries. Not only surgeons, but also patients can benefit from the pre-operative outcomes of breast shape prediction. The parametric model can play important role to improve the patient’s quality of life by filling the gap between pre and post-operative outcomes.

In terms of hardware devices, common 3D reconstruction sensors require expert personnel. It is essential to propose such method which can be performed even without skilled staffs. Regarding the cost, the proposed method would be equipped with low-cost devices that are accessible by all clinics. Moreover, it is aimed to perform more accurate reconstruction to result models appropriate for medical diagnosis applications.

6 Research Milestone

Future research can be concentrated on employing different methods to reconstruct the parametric model of the breast. Considering both the advantages and disadvantages of previously researches, we tend to introduce a method which can be used in breast surgery planning and prediction.

Methods such as [15] used low-cost active sensors to perform the reconstruction; however they have not modeled the breast completely. Continuing the research done in [15], we aim to perform the complete 3D reconstruction by utilizing low-cost devices. Keeping the cost in mind, dissimilar to methods such as [14], we will propose a complete automatic method in order to ease the process of scanning for non-expert personnel. The novelty of current research is to perform complete 3D reconstruction together with planning and prediction which will be done automatically.

The trend is continued to relate the model to parameters which control the shape of the breast. Rigid model is converted to alive and elastic model which is altered based on any tissue removal. Although parametric reconstruction has been performed in some parts of human body in previous works, it has not been proposed on breast yet, thus we tend to specify breast characteristics in the reconstructed model. Mentioned parameters provide the facility for the surgeon to track the breast shape changes as any part of tissue is removed.

Computation complexity is an important factor to reduce the cost. The proposed method in either reconstruction or parameterization requires common hardware which can be obtained without significant cost.

Finally study would be resulted in a planning application to be applied by surgeons and specialists. Both planning and prediction are considered to assist the surgeons in the package. The patient’s breast is scanned with appropriate scanning device. It will then be modeled in reasonable time. Proposing a user-friendly application is the second aim of the current task to ease the interaction between the surgeon and specialists with the computer.

Acknowledgment

This work is financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project 3D Models for Aesthetic Evaluation and Prediction of Breast Cancer Interventions with reference PTDC/SAU-ENB/114951/2009.

References


Total Variation Denoising using a Recursive and Spatially Adaptive Filter

João M. R. Sanches
 Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisbon, Portugal

Abstract

This paper presents a method for solving the total variation (TV) regularized denoising problem for data fidelity terms corresponding to different observation and noise models - additive Gaussian, multiplicative and Rayleigh, and Poisson. The TV term is reformulated as a weighted sum of difference terms and the data fidelity term is reformulated using a Taylor series approximation into a weighted least squares term for which only some parameters depend on the model. The maximum a posteriori probability (MAP) solution is computed by applying a recursive and spatially adaptive filter on the maximum likelihood estimate, with the weights corresponding to the TV term updated recursively.

1 Introduction

Image denoising/reconstruction involves estimating an image or three-dimensional (3D) volume from a set of noisy and often incomplete set of observations. The relationship between the image to be estimated, \( x \) and the observation, \( y \), and the type of noise corrupting the observed image depends on the imaging modality. Since estimating the image from the noisy observations is an ill-posed problem, we need to make some assumptions about the characteristics of \( x \), which are incorporated in the prior in the Bayesian Maximum A Posteriori Probability (MAP) inference framework, and the regularizer term in the related convex optimization formulation. Total Variation (TV) [9] has emerged as a popular choice of regularizer, to promote piece-wise smoothness in the estimate. The isotropic TV semi-norm of a two-dimensional (2D) image \( x \) indexed by coordinates \((m,n)\) (assuming suitable boundary conditions) is defined as,

\[
TV(x) = \sum_{(m,n)} \left( (x_{m,n} - x_{m-1,n})^2 + (x_{m,n} - x_{m,n-1})^2 \right)^{1/2}.
\]

(1)

A generalized expression for (1) which can be extended to 3D is,

\[
TV(x) = \sum_i \left( \sum_{j \in N_i} (x_j - x_i)^2 \right)^{1/2},
\]

(2)

where the index \( i \) indexes the image or volume, and \( N_i \) is its neighborhood, over which the difference is computed. Unless stated otherwise, the neighborhood \( N_i \) is the first order neighborhood. For the 2D case the index vector is \( i = (m,n) \) and the neighborhood is \( N_i = \{ m-1, m, m+1 \} \times \{ n-1, n, n+1 \} \).

For a given statistical model, the MAP estimation with TV regularization is equivalent to solving the optimization problem,

\[
\min_x J(x,y) + \frac{\lambda}{2} TV(x),
\]

(3)

where \( J(x,y) \) is the data fidelity term and \( \lambda > 0 \) is the regularization parameter. Refer to table 1 for the data fidelity term, likelihood \( p(y|x) \), and maximum likelihood (ML) estimate \( x^{ML} \) for the Gaussian, Rayleigh, and Poisson models.

<table>
<thead>
<tr>
<th>Noise Model</th>
<th>Gaussian</th>
<th>Rayleigh</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(y</td>
<td>x) )</td>
<td>( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} )</td>
<td>( \frac{1}{\mu} e^{-</td>
</tr>
<tr>
<td>( l(x,y) )</td>
<td>( \frac{1}{2} \sum_{i} (y_i - x_i)^2 )</td>
<td>( \frac{1}{2} \sum_{i} -\log(x_i) + \frac{y_i^2}{x_i^2} )</td>
<td>( \frac{1}{2} \sum_{i} -\log(y_i) + \frac{x_i^2}{y_i^2} )</td>
</tr>
<tr>
<td>( x^{ML} )</td>
<td>( y )</td>
<td>( \frac{y}{\mu} )</td>
<td>( \frac{\gamma}{\mu} )</td>
</tr>
</tbody>
</table>

Table 1: Parameters for the noise models.

TV regularization has been widely used in denoising with additive and Gaussian noise [3, 7, 13], and also in the context of ultrasound de-speckling [11], SAR denoising [2], and poisson noise [8]. However, the methods for non-Gaussian noise are application specific and may not be useful as a proximal mapping [4] for the respective noise model.

1.1 Contributions

In this paper we propose a method for TV denoising which only needs some parameters initialized depending on the observation model. We extend the approach for a quadratic Gibbs prior from [10] to the more general TV regularizer. The MAP estimation is formulated as a filtering operation on the ML estimate, through a Taylor series approximation and solving for the stationary condition. Because of the non-smooth nature of the TV term as opposed to the quadratic prior, the filter does not have a closed form. This necessitates a recursive approach in the iterative reweighted least squares (IRLS) framework [5] in which the filter weights are updated after each iteration. Because of the dependence on the neighboring pixels, the recursive filter is spatially adaptive unlike IRLS for the \( \ell_p \), \( 0 \leq p \leq 1 \) norms.

2 Proposed Method

The ML estimate of \( x \), given the observation \( y \) is the value of \( x \) which minimizes the negative of the log likelihood function, \( x^{ML} = \min_x l(x,y) \). Since the data fidelity term \( l(x,y) \) is separable, we can solve for each element of \( x \) of \( y \) using a Gauss-Seidel approach. Each element is estimated by solving,

\[
\frac{\partial l(y,x)}{\partial x_i} = 0.
\]

(4)

We first approximate the likelihood function by the second order Taylor series, computed at the ML estimate, \( x^{ML} \), i.e.,

\[
l(y,x) = l(y,x^{ML}) + \frac{\partial l}{\partial x_i} (y,x^{ML})(x_i - x_i^{ML}) + \frac{1}{2} \frac{\partial^2 l}{\partial x_i^2} (y,x^{ML})(x_i - x_i^{ML})^2.
\]

(5)

As done in the IRLS method for non-smooth \( \ell_p \), \( 0 \leq p \leq 1 \) norms, we can express the TV term in terms of reweighted quadratic terms,

\[
TV(x) = \sum_i \sum_{j \in N_i} (x_j - x_i)^2 = \sum_i \omega_i \sum_{j \in N_i} (x_j - x_i)^2
\]

(6)

where the coefficient at location \( i \) is, \( \omega_i = \left( \sqrt{\sum_{j \in N_i} (x_j - x_i)^2 + \epsilon} \right)^{-1} \), with \( \epsilon > 0 \) a bias term to prevent the weight from becoming very large in regions of small difference between neighboring pixels.

As in [10], replacing (5) and (6) in (3), the computation of (4) leads to, after gathering all terms which contain \( x_i \),

\[
-\frac{\partial^2 l}{\partial x_i^2} (y,x^{ML})(x_i - x_i^{ML}) + 2\lambda \omega_i \sum_{j \in N_i} (x_j - x_i) + 2\lambda \sum_{k \in N_i} w_k (x_i - x_k) = 0,
\]

where \( N_i \) is the set of all neighbors of the \( i \)-th node, for which \( x_i \) is involved in computing the difference at that node. The term \( \frac{\partial^2 l}{\partial x_i^2} (y,x^{ML}) \) in (5) is zero by definition, because the ML estimate \( x^{ML} \) is a stationary point of the likelihood function. Hence, we have two first-order difference terms in (7), corresponding to the weight \( w_i \) of the pixel \( i \), and to the sum of difference terms involving pixel \( i \).

Replacing \( \gamma = \frac{\partial^2 l}{\partial x_i^2} (y,x^{ML}) \) in (7), we get the first-order equation,

\[
y(x_i - x_i^{ML}) + 2\lambda w_i \sum_{j \in N_i} (x_j - x_i) + 2\lambda \sum_{k \in N_i'} w_k (x_i - x_k) = 0.
\]

(7)
RecPad 2013

\[
\hat{x}_k = \frac{\gamma + 2\lambda w_iy_k + 2\lambda \sum_{k \in N_j} w_k}{\gamma + 2\lambda w_iy_k + 2\lambda \sum_{k \in N_j} w_k},
\]

where \(y_k\) is the size of the neighborhood. This leads to a spatially adaptive filter, with the closed form,

\[
x_i = a_i \hat{x}^{ML}_i + b_i \sum_{j \in N_i} x_j + c_i \sum_{k \in N_i} w_k x_k,
\]

where \(a_i = \gamma/d_i, b_i = 2\lambda w_i/d_i, c_i = 2\lambda/d_i\), with the denominator term \(d_i = (\gamma + 2\lambda w_i y_k + 2\lambda \sum_{k \in N_j} w_k)\). The weights \(w_i\) at a pixel \(i\) and the filter coefficients \(a_i, b_i, c_i\) are functions of the MAP estimate \(x\). Equation (9) can therefore be expressed in a matrix form,

\[
x = A(x) x^{ML} + (B(x) + C(x)) x,
\]

which can be expressed in terms of a filtering operation on \(x^{ML}\),

\[
x = (I - B(x) - C(x))^{-1} A(x) x^{ML}.
\]

However, since the filter term in the right hand side of (11) is a function of the MAP estimate \(x\), we need to use a recursive approach to estimate \(x\) and update the matrices \(A(x), B(x), C(x)\). From (9), we express the estimate at iteration \(t + 1\) in terms of the recursive filter, in matrix notation as,

\[
x^{t+1} = A(x^t) x^{ML} + (B(x^t) + C(x^t)) x^t.
\]

It can be shown that the proposed recursive filter is a Majorization-Minimization (MM) algorithm, for which convergence conditions can be verified [12].

3 Experimental Results

Our method was implemented in Matlab, with the spatially adaptive filter implemented in C++ through Mex. All experiments were performed on an Ubuntu Linux laptop with an Intel i5. In table 2, we compare our proposed method with other solvers for TV denoising for additive and Gaussian, multiplicative and Rayleigh, and Poisson noise, on the size 128 \times 128 Shepp-Logan phantom. The measure of error used was the mean square error (MSE) for Gaussian noise, and the normalized Mean Absolute Error (MAE) for the Rayleigh and Gaussian noises. The noisy images and estimates are presented in figure 3. We see that our method leads to the lowest error in each case.

Table 2: Comparison of TV filtering for Gaussian, Rayleigh, and Poisson noise, with other solvers for TV denoising.

<table>
<thead>
<tr>
<th>Model</th>
<th>Method</th>
<th>iterations</th>
<th>CPU time (sec)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Proposed</td>
<td>309</td>
<td>2.7</td>
<td>0.0041</td>
</tr>
<tr>
<td></td>
<td>Chambolle [3]</td>
<td>33</td>
<td>0.33</td>
<td>0.0187</td>
</tr>
<tr>
<td></td>
<td>Split Bregman [7]</td>
<td>-</td>
<td>0.01</td>
<td>0.0073</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>Proposed</td>
<td>19</td>
<td>0.33</td>
<td>2.535e-05</td>
</tr>
<tr>
<td></td>
<td>ADMM [1]</td>
<td>59</td>
<td>2.47</td>
<td>2.55e-05</td>
</tr>
<tr>
<td></td>
<td>LPAMA-mV [14]</td>
<td>96</td>
<td>0.107</td>
<td>6.039e-05</td>
</tr>
<tr>
<td>Poisson</td>
<td>Proposed</td>
<td>16</td>
<td>0.13</td>
<td>1.345e-05</td>
</tr>
<tr>
<td></td>
<td>PIDAL [6]</td>
<td>200</td>
<td>23.72</td>
<td>1.757e-05</td>
</tr>
<tr>
<td></td>
<td>Log TV [8]</td>
<td>1</td>
<td>0.01</td>
<td>3.219e-05</td>
</tr>
</tbody>
</table>

4 Conclusions

We have proposed a method for TV denoising through recursive spatially adaptive filtering. The proposed method can be applied to different observation models, and can be easily extended to the higher order total variation. Optimizing scanning patterns across the image or volume for better computational efficiency is a possible direction for future research.

Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia (FCT), Portuguese Ministry of Science and Higher Education, through a Post-doctoral fellowship (contract no. SFRH/BPD/79011/2011) and FCT project PTDC/EEP-EEL/024187/2011/2009/2013).

References


Figure 1: Shepp-Logan phantom: (a) original, (b) image with Gaussian noise (SNR 0 dB), (c) estimate using the proposed method, (d) image with multiplicative Rayleigh noise, (e) estimate, (f) image with Poisson noise, (g) estimate.
Selection of epilepsy-related EEG ICA components for simultaneous fMRI analysis

Rodolfo Abreu¹
rodolfo.abreu@ist.utl.pt
Alberto Leal²
a.leal@aleeg.com
Patrícia Figueiredo¹
patricia.figueiredo@ist.utl.pt

¹ Institute for Systems and Robotics / Instituto Superior Técnico
Universidade de Lisboa
Lisbon, PT
² Centro Hospital Psiquiátrico de Lisboa
Lisbon, PT

1 Introduction
With the possibility of recording EEG and fMRI simultaneously, the number of research studies trying to correlate the electrical activity recorded by the EEG with the blood-oxygen-level-dependent (BOLD) signal has significantly increased, particularly in epilepsy where the mapping of the epileptogenic focus (or foci) is the main goal [5].

When integrating EEG and fMRI data simultaneously, spatio-temporal filtering must be applied to the EEG in order to extract a representative time-course of the activity of interest. The most common approach resides in using ICA. This method tries to decompose the data into a set of statistically independent components that are described by their time-courses and corresponding spatial maps.

Subsequently, the challenge of selecting these components of interest rises since one is interested in keeping the meaningful components for further analysis while removing the remaining ones. The classical approach usually resides in manually selecting the components of interest based on a set of criteria defined and assessed by the researcher. Some authors state that the selection steps should be firstly focused on the spatial distribution, searching for topographies consistent with the expected topographical activation patterns [3]. In contrast, other authors look at the temporal dynamics of the components time-courses, seeking for the ones that carried the activity of interest [3].

The ICA can also be used to separate brain processes from non-brain ones and some approaches focused on automatically selecting the artefactual components in order to obtain an artefact-free EEG [1]. To our knowledge, only one study tried to automatically select independent components of interest for an epilepsy-unrelated Event-Related Potentials (ERP) study [6]. The authors implemented an algorithm – COMPASS – that performs a components pre-selection based on the scalp distribution of the topographies, also finding the ERP-related channels. The final selection is done by means of time-correlation between the pre-selected components time-courses and the ERP-channels.

2 Materials and Methods

2.1 EEG-fMRI Data Acquisition and Processing
Two paediatric (GB, BS) and one adult (PS) patients with drugs-refractory focal epilepsy undergoing pre-surgical evaluation were selected from the Program of Surgery for Epilepsy of the Hospital Center of West Lisbon. All patients were studied on a 3T, 12-channel RF coil, Siemens Verio scanner using an MR-compatible 32-channel EEG system (Brain Products). For all paediatric patients, two fMRI runs (10 and 20 minutes) were collected using a gradient-echo echoplanar imaging (EPI) sequence (TR/TE = 2500/30 ms, 3.5 × 3.5 × 3.0 mm³ voxel size). Structural images were collected using the MPRAGE sequence. For the patient PS (with an epileptogenic cortical dysplasia near the primary motor cortex), three 10min fMRI runs and two additional 5min finger-tapping motor-task fMRI runs were collected, allowing the mapping of eloquent motor cortex, better informing the pre-surgical planning. For cross-validation of the intra-MR EEG traces, a 10min EEG recording was also performed outside the scanner.

Standard EEG pre-processing steps were applied to the intra-MR EEG recordings: gradient and pulse artefacts reduction algorithms were applied, downsampled to 250Hz and band-pass filtered (1 – 45 Hz). The interictal activity was visually inspected by an neurophysiologist (AL). Next, the Infomax ICA algorithm as implemented in the EEGLAB toolbox was applied and a set of epilepsy-related components were selected by running our algorithm. Metrics were then extracted from the selected components time-courses, convolved with the canonical haemodynamic response function (HRF), downsampled to the fMRI sampling rate and used as a regressor of interest in a general linear model (GLM) analysis of the fMRI data using FSL. Finally, Z-statistic maps were obtained and cluster-thresholded with Z > 2.3 and a cluster significance threshold p = 0.05 [5].

2.2 Selection of Independent Components

The epilepsy-related components were selected using the concept of template of the epileptic activity. Since the ICA algorithm returns two sources of information about the components, dataset-specific temporal (spike) and spatial (topography) templates were computed.

**Spike Template.** After a standard events alignment step, the EEG data is epoched obtaining a N × L × E matrix, where N represents the number of channels, L the length of each epoch and E the number of epochs (in our case, spikes). In order to find a spike template, the mean epochs across trials was computed, obtaining the matrix $M = (m_1, \ldots, m_N)$. The mean epoch that will be used as template is given by:

$$SP = M(n) = \max_n \{VAR[M]\}$$

where $VAR[M]$ represents the variance of $M$ and $n = 1, \ldots, N$.

RecPad 2013 Page 89 of 147
Topography Template. In order to find the instant in which the topography template will be obtained, the first peak of the Global Field Power (GFP) is computed for the matrix $M$. The GFP represents a measure of the spatial standard deviation. If the first peak of the GFP is located at the instant $t$, the topography template will be given by:

$$ T = (m_1(i), \ldots, m_N(i)) $$

(2)

Features extraction and Components selection. Once both spike and topography templates are built, correlation measures between them and the components time-courses and topographies, respectively, are computed. Taking into account possible deviations for the spikes latencies in the IC-space, a shift $b (b \in [-0.040, 0.040] \, \text{ms})$ was added and the time-correlation between $SP$ and each epoch was computed for each value of $b$. The maximum correlation within the values of $b$, $C_{\text{max}}$, is found and stored in a $N \times 4$ matrix $C$, containing all the values of $C_{\text{max}}$ for each trial of each channel. The mean correlation across trials is computed, normalized to its maximum and squared to increase the weight of the high correlation values, obtaining the array $MC_{\text{norm}}$.

In order to compute the correlation between the topography template and the components topographies, the concept of spatial correlation is used, which is defined by:

$$ SC(n) = \frac{\sum_{i=1}^{N} T_i \times t_i(n)}{||T|| \times ||t(n)||} $$

(3)

where $T$ and $t(n)$ denote the mean-subtracted and GFP-normalized topography template and the $n$th component topography, respectively.

Concatenating both features, an $2N$ array, $C_{\text{norm}}$, is built. For the clustering procedure, we restricted the number of clusters to 3 or 4 and, for each case, the mean silhouette was computed to select the number of clusters that more accurately divided the data into $k$ partitions. After running the $k$-means clustering algorithm over $C_{\text{norm}}$, another array, $CL$, containing the clusters which each observation of $C_{\text{norm}}$ will be assigned to is obtained. Finally, our algorithm reshapes $CL$ into a $2 \times N$ matrix, $CL'$, where $CL'(1)$ and $CL'(2)$ contain the clusters that the observations within $MC_{\text{norm}}$ and $SC$, respectively, were assigned to. Being $c$ the cluster described by the highest temporal centroid, the $i^{th}$ component is selected if the following condition is verified:

$$ CL'(1,i) = CL'(2,i) = c $$

(4)

Using this methodology, the components with the highest temporal and spatial correlations which are also consistent with each other are found. Due to employment of clustering routines, no restrictions regarding the number of components to be selected are placed, enabling the selection of more than one meaningful component.

3 Results and Discussion

In Figure 1 seven topographies and a 10 seconds trace from the component #1 are shown. For this dataset, the components #1, #2 and #6 were selected for further analysis. The component #7 was also selected by the COMPASS algorithm although no epileptiform activity was found. This is probably due to pre-selection step based on the spatial correlation, which is quite high for this component, rendering flawless decisions.

Figure 1: Spatial and temporal correlations. Seven component topographies and a 10 seconds trace from component #1 are shown. The dashed black squares indicate the epilepsy-related components that were selected. The dashed red square indicates the additional component selected by the COMPASS algorithm. The spikes are represented by the red vertical lines.

In Table 1 a comparison between the components that were manually and automatically selected in each dataset is presented. Overall, our algorithm performed better than the COMPASS in selecting epilepsy-related components. These results were further validated by the identification of plausible epileptic networks in the fMRI data.

Table 1: Comparison between the components that were automatically selected by both methods with the ones selected by the neurophysiologist.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Manual</th>
<th>Automatic</th>
<th>COMPASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB</td>
<td>1, 2, 6, 12</td>
<td>1, 2, 6, 7, 13</td>
<td></td>
</tr>
<tr>
<td>BS</td>
<td>2, 5, 8</td>
<td>5, 8, 15, 18</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>1, 4, 4, 10</td>
<td>3, 3, 3, 13</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Patient GB: Epileptic focus mapping. From left to right: Coronal, sagittal and axial structural slices. Brain regions with epilepsy-correlated BOLD changes (in red-yellow) are superimposed. The intersection of the axis points to the voxel with maximum z-score.

4 Conclusion

In this paper, a data-driven algorithm able to automatically select epilepsy-related components was developed, making it an important contribution in simultaneous EEG-fMRI epilepsy studies by providing an user-independent and objective alternative for selecting meaningful components.

This work was supported by the FCT project PTDC/SAU-ENB/112294/2009.

References

Detection, classification and localisation of football players and ball from Handycam videos

Tiago Vilas
tiago.silva.33@hotmail.com
J.M.F. Rodrigues
http://www.ua.pt/~jrodrig
Pedro Cardoso
http://www.ua.pt/~pcardoso

Abstract

The current performance requirements in football make imperative the use of new technologies for game observation and analysis, providing detailed information about the team’s actions. This paper presents in resume an algorithm to collect information from Handycam videos, namely, the classification within the team, the tracking and localisation of players and ball, being the main contribution the determination of the exact localisation of players and ball in the pitch.

1 Introduction

Footdata [1] is a project to build a new multi-platform product for football. Footdata is based on information technologies, and integrates two fundamental components of the football world: i) a social network, with all the typical features and ii) the professional component, which features an acquisition and information system to meet all the football management needs. One of the modules is a framework that reports the exact positions of the players and ball in the football pitch.

There are several models available in the literature do detect and track players and ball, e.g. [2-4]. In this paper we present in resume an algorithm that can be applied to Handycam videos that allows: (a) the detection of the football pitch, (b) the detection of players and ball, (c) the assignment of players to their teams, (d) the tracking of players and ball and (e) their localisation (in meters) in the pitch. The main contribution is the exact localisation of players and ball from Handycam videos.

2 Football pitch detection

Let $I_{GB}(x,y,t)$ and $I_{HSV}(x,y,t)$ represent a frame with dimensions $M \times N$, acquired with an Handycam (FullHD or HD) situated for instance in the stands, where $M$ is the width and $N$ is the height of the image, $t$ is the instance when the frame was acquired and RGB and HSV the colour spaces.

The first step was the football pitch detection, which consists in (a) the segmentation of the green grass regions for each frame using $I_{HSV}$. This was done by a semi-automatic process, where selected pitch patches were used to compute the HSV thresholds intervals. For the pitch shown in Fig. 1, it was used the following thresholds levels: $H \in [0\% , 50\% ]$, $S \in [12\% , 100\% ]$ and $V \in [27\% , 100\% ]$, which returns a binary image $I_c$, with $I_c = 1$ corresponds to the green pixels.

The next step has the purpose to eliminate the areas outside the pitch. This was carried out in several stages: First (b), in $I_c$ was chosen a central point of the image as the seed (in most frames, the central area corresponds to the pitch, the exceptions are zooms to players or to the stands). The central point of $I_c$ has to have a green colour (see above), and recursively from this initial point, using $3 \times 3$ neighbourhood, all central pixels that have all neighbours green are kept, $I_c' = 1$, the remaining are put to 0 ($I_c' = 0$).

Over the last image, but only in areas where exist pitch, i.e. $I_c' = 1$, was applied (c) the Canny edge ($I_{CE}$) detection [5] on $I_{GB}$. After this, (d) the Hough transform [5] was applied over the $I_{CE}$ in order to detect the lines ($I_{H}$) that limits the pitch. Finally (e), using $I_{H}$ the most horizontal and vertical lines (left, right, up and down) are computed.

Those lines correspond to the limits of the pitch, and everything outside these limits (lines) was removed. The final results consist only in the segmentation of the pitch (with some black blobs inside), $I_{fj}$.

3 Players and ball detection

For the player detection it was used the $I_{fj}$ image. In this image, most of the work for the players’ detection was already done, once the pitch area was delimited, the black blobs inside this area are considered players. Nevertheless, 3 major problems arise: (a) blobs that are not players, e.g., areas with short grass, (b) players that almost disappear due to low image definition, or been partially equipped in “green,” and (c) the overlap of various players.

To solve these situations and to obtain only the contours of the players, morphological filters were applied, i.e., first it was applied the dilation filter (D) followed by an erosion filter (E) and finally a $9 \times 9$ medium filter (M). The result was an image where the players were better defined and specific noise regions were removed, $I_{fj}(x,y,t) = M(E(D(I_{fj}(x,y,t))))$.

After this, some player still have more than 1 region (usually due to the equipment that they use), a vertical mask was applied to connect those regions. The next step was the confirmation that each blob corresponding to a possible player (in $I_{fj}$) has at least one edge contour correspondence in $I_{CE}$. Finally, it was checked if the final blob region of each player (in $I_{fj}$) doesn’t have more than 10% of green pixels (using $I_{H}$). The final image after this validation was called $I_f$.

To differentiate between players and the ball, all pixels with the “white colour” (in $I_{fj}$) inside $I_f$ were detected. Those pixels can represent lines belonging to the pitch, the ball, and if it was the case, teams that had white predominantly equipment’s. For this purpose an threshold with $V \in [75\% , 100\% ]$ was applied to $I_f$, returning a binary image, $I_b$, with all the white regions inside the pitch ($I_b = 1$).

Combining the previous information with the size and shape of the blobs, the blobs that corresponds to the balls can be classified. The first step, consists in detect if the blob as more or less circular or oval shape, which was different from the player blobs that had (usually) a more “vertical-rectangular” shape. This way most of the players blobs were discarded. The second step, was to remove the areas corresponding to the pitch marks, eliminating the areas that correspond to the lines detected by the Hough transform in $I_{H}$.

The final step consists in removing areas which exceed $M \times N/k$ pixels ($k = 10000$, this value was empirically determined).

Figure 1 top shows the player delimited by rectangles, and the ball by a circle. Having the ball classified, it was necessary to classify the blobs as players from team A or B, referee or keepers.

4 Assignment of the players to their teams

The players’ classification to a team was based on the colour of the equipment. It was assumed that the video starts at the beginning of the game (with the ball in the middle of the pitch), if that doesn’t happen, then a semi-automatic process selects the most probable 7 player from each team. Considering that the video starts at the beginning of the game, then a group of 7 players on the right, and 7 players on the left of the middle pitch line (represents each team) was automatically selected.

For each team, in separate, and using the blobs regions defined as players in $I_f$, the average of Hue using the $I_{H}$ image was computed. Having the two average Hues values from team A and B, the middle threshold was computed to separate the teams.

The keepers were relatively easier to classify, once they correspond to the first blob that appear in the right and in the left of the image with significant difference from the two average Hues values computed for the teams (it is important to remember that the keepers use a different equipment from the team). The major problem was the referee, they usually use black, but that is not mandatory. If the video starts from the beginning of the game, the referee usually was very close to the middle line, in those cases it was a blob with a Hue different from the players near the middle line. If the video starts from a different position then a semi-automatic process is applied, i.e., the referee was considered the blob with the most different Hue (comparing with the Hues from the teams), and if by mistake was assigned as a team player, then a web base tool (out of the focus of this paper) allows the user to manually change the player classified as referee and vice-versa.

All those Hues once detected were memorised by the system and adapted dynamically in function of the video conditions (light, etc.).

The numbers assigned to the players were in function of the position they occupy in the pitch, and those number were associated to the real player number and name in function of the position posted in the game.
sheet (e.g. right defender). If necessary, this can be corrected later using the web base tool mentioned above.

Finally, obviously there is also the validation that it was impossible to assign more than 1 referee, 2 keepers and 10 players per team. Figure 1 shows the result of player’s assignment to a team and referee.

5 Players position in the pitch

In order to make a correct analysis of a soccer game, it is important to have knowledge of the correct position of the player in the pitch. To calculate that position, a perspective transformation from the frame to a normalized pitch was needed, see Fig. 1 bottom.

For this purpose a set of references in the pitch was computed. Those were extracted from the lines detected in the Hough transform, \( \text{Hough} \) and the pitch delimitations in \( I_f \). For the position of the player, it was considered the middle point in the bottom line that limits the player, which corresponds in most cases to the coordinates of the players’ feet, see Fig. 1 top and the respective projection in the bottom.

Since the pitches have different sizes (weight and height), it was important to compute the size of a pixel (in our case, in meters). For this computation, in the first frames (as soon as possible) the lines of the centre circle, penalty area or goal area has to be automatically detected to calibrate the value of a pixel in meters (the dimensions of these areas are the same for all pitches). An automatic process was then applied every time there was a small zoom, or a zoom to a player.

6 Tracking of players and ball

The players and ball tracking avoids doing all the steps mentioned in Section 3 and 4 for all frames. The implemented tracking process, though simple, was effective except in situation of great confluence of player, such as corner, or discussions with the referee. Nevertheless the main focus of this framework is not to follow the players’ everywhere, but to compute the distances between players and players to the ball.

The tracking was divided into five steps for each frame \( t \): (a) the distance (in meters) was calculated for each player, between the previous frame \( t-1 \) to the current frame \( t \). Making the correspondence to the player that was closer. If there were two or more players at the same distance (non-overlapping) the player was assigned to the one that in the previous frame has the nearest Hue.

(b) If there was no player in frame \( t \) near to the position occupied in frame \( t-1 \) (few meters apart), then it was checked in frame \( t \) with increasing circle diameters the player with the same colour as the player in frame \( t-1 \). It was used the trajectory of the previous 5 frames to limit within the circle the “sectors” of this search.

(c) For situations in which a particular player was not tracked/located during a long frameset (2 seconds), this player was then searched from the position where he was lost, with increasing circle diameters, for shots not tracked and has the same Hue. The first shot found with this characteristic was assign as the lost player.

(d) It is important to note that in the majority of the frames, there was the possibility that some players were not detected because they are not in the field of view of the camera, e.g., at the beginning of the game, due to fact that the camera doesn’t capture the entire pitch. Every player that appear at the left or right (top or bottom, depending if there has a zoom applied) was considered as new player to track, except if a player was “lost” on a previous frame in a nearby position.

(e) Finally, it was also necessary to verify the existence of collisions. If the above procedure (a-c), a player was not found, it was assumed that there was a collision (junction of two or more players). For those situations, all shots that correspond to players, that were in positions very close and were lost, were associated with a single shot, assigning to 2 or more players that were tracked in \( t-1 \) in a nearby position. When the collision ends (separation of two or more players) the Hue of each shot were checked and players reassigned.

After a collision there was always the possibility of players being wrongly identified, for instance when the collision was between two players of the same team. To solve this, the already mentioned web base tool can be used to correct the numbers of the players or even the team. The process of tracking the ball and referee is similar to the players.

7 Discussion

This paper presents a proposal for detection and tracking of football players and ball with the objective of computing distances between players and players to the ball. For now, it was an initial framework that still needs improvements. Nevertheless some tests were carried out showing promising results.

Five different games, with approximately 15 minutes of each video (4 FULL HD and 1 HD) were analysed. The results obtained were: (a) the detection of players with 92.3% of correctness, (b) 2.0% of assignments of detected players to the wrong team and (c) 0.4% of false positives, i.e., detection of players, in locations where there weren’t. To compute this numbers all the situations where a cluster of players occurs were removed, such as corners and discussions with the referee, etc.

It is also important to mention, that in cases when a player is expelled or when there is a substitution it is necessary to pass this information using the web tool.

All the steps presented can be subject to numerous improvements, one essential was the code optimization in terms of time consuming. Despite the code has already some parallelisation, an entire or partial implementation using the GPUs is expected to improve times.

Figure 1: Top, the result of the ball, referee and player’s detection and assignment to team. Bottom, the representation of the players position considering the origin of coordinates the interception between the middle line and the top lateral line.

Acknowledgments

This work was supported by FCT project PEAT-GE/EEI/LA0009/2013 and project FootData QREN I&D T, n.º 23119. We also thanks to project leader Inesting, S.A. [www.inesting.com], and the consultant football coach Domingos Paciência and our colleague Carlos Gomes.

References


Mosaicing the Interior of Tubular Structures

D. Pereira, J. Tomaz, R. Ferreira, J. Gaspar
http://www.isr.ist.utl.pt

Institute for Systems and Robotics / IST
Lisbon, Portugal

Abstract

This paper addresses a Simultaneous Mapping and Localization SLAM methodology for a system capable of performing visual inspections in an unknown environment domain assumed to be a tubular shaped structure (TSS), using a monocular camera. Under the TSS assumption, a geometrical model was developed, which directly maps visual features from planar images onto a 3D representation, through a backprojection procedure, thus enabling a scenario reconstruction composed of cylindrical segments. The Extended Kalman Filter (EKF) allows handling non-linear systems and measurement models with zero-mean multivariate Gaussian distributed noise. The EKF framework allows reconstructing the path described by the camera and its visualized structure, based in visual landmarks described by feature detectors.

1 Introduction

Estimating the motion of a camera moving inside a Tubular Shaped Structure (TSS) involves considering various aspects. Two important aspects are the number of degrees of freedom of the camera motion and the shape of the environment structure. In this paper, the TSS involves straight and curved sections, allowing free movement of the camera. Motion can be estimated by registering the texture, retrieving distinctive visual features, and dewarping into a mosaic. Hence we start from the standard idea of reconstructing points of the scene and then focus on fitting a simple 3D cylindrical model to the various tube sections, which makes simple the dewarping step [4].

2 Estimation of Tubular Structure and Camera Motion

Assuming the world as a TSS domain navigated by a forward hand-held like monocular camera (see Fig 3), smoothly moving with a constant speed inside a tube without revisiting previous positions, one can determine the features 3D locations, which are structured to a cylindrical section with radius $\rho$. By thoroughly defining a state vector $s_t$ composed of both cameras and the TSS geometrical parameters estimates, and incorporating a geometrical observation model reliable enough to produce relevant estimations of the cylindrical section geometrical parameters, the simultaneous camera localization and mapping can be achieved.

Figure 3: TSS Model (left) and tube parameters (right).

2.1 Extended Kalman Filter

In this section we define a filtering methodology that allows estimating both the TSS shape and the camera motion. The first step consists of defining the state vector of the filter, $s$, as a joint composition of camera pose $x$ and cylinders geometrical parameters $y$:

$$ s_t = [x_t; y_t] $$

where the semicolon denotes vertical stacking of vectors.

Camera motion behavior is modeled with a constant velocity dynamic model. The state vector $x_t$ provides the position $r^C_t$, orientation $\theta^C_t$ and both linear $v^W_t$ and angular $\omega^W_t$ velocities at every instant, i.e.

$$ x_t = [r^C_t; \theta^C_t; v^W_t; \omega^W_t]. $$

A cylindrical section is characterized by a 7 dimension $y_{nk}$ state vector as an array of the cylinders parameter centres position $p_{nk}^W$, orientation $o_{nk}$, and radius log($\rho_{nk}$) expressed in a logarithm form

$$ y_{nk} = \left[ p_{nk}^W; o_{nk}; \log(\rho_{nk}) \right]. $$

A Kalman Filter estimates optimal states over time, given observations in the presence of noise of a dynamic system driven by noise inputs. This estimation is computed recursively as a Markov chain model, i.e. it only requires the previous estimate $k - 1$ of a given state at instance $k$. An assumption required by this framework is to use linear systems and observations models with zero-mean multivariate Gaussian distributed noise.

2.2 System Dynamics

The non-linear state transition model function $f$, is defined as the stacking of two independent state transition processes, $f_1$ and $f_2$, for both camera and cylinder state vector $x$ and $y$, which are influenced by an additive zero-mean Gaussian transition noise $\xi \sim \mathcal{N}(0, Q_k)$:

$$ f = [f_1; f_2]. $$

The constant velocity model allows smooth velocity variations with zero-mean Gaussian distributed acceleration noise $n_k \sim \mathcal{N}(0, N_k)$. The acceleration noise has linear and angular components $n_k = [a^W; \omega^C]$. Under this model assumption, and defining $V^W = a^W \Delta t$ and $\Omega^C = \omega^C \Delta t$ as the linear and angular velocities impulse between a transition step $\Delta t$, one can express the camera state transition, $x_{k+1} = f_1(x_k, n_k)$, 1, 2:

$$ \begin{bmatrix} r^W_{k+1} \\ \theta^C_{k+1} \\ v^W_{k+1} \\ \omega^W_{k+1} \end{bmatrix} = \begin{bmatrix} r^W_k + (\frac{V^W_k}{2} + V^W_{k+1}) \Delta t \\ \theta^C_k + \omega^C_k \Delta t \\ v^W_k + V^W_k \Delta t \\ \omega^W_k + \Omega^C_k \Delta t \end{bmatrix}. $$

2.3 Observation Model

In this work, we use the pin-hole model to describe geometrically the 2-D imaging of a 3-D point:

$$ \tilde{m} \sim \mathcal{N}(M, P) $$

where $M = [X Y Z]^T$ denotes a 3D point, $\tilde{m} = [\mu \nu]^T$ is the respective image, and $P$ is the projection matrix [3]. Conversely, to describe the 3-D location of points $M$ in space, imagined as $m$ in the image plane, one can relate the 3-D ray leaving the camera optical centre $C$ and piercing the image plane at $m$ as a line which intersects the plane where $M$ is lying. $C$ and $D$ define a line in 3-D space, where every point in the line is thus given by the following equation $M = C + \alpha D$, with a scaling factor $\alpha \in [-\infty, +\infty]$. The full expression for representing a point $m$ back-projection into $M$ is

$$ M = -p_{(123)}^{-1} p_{(4)} + \alpha p_{(123)}^{-1} \tilde{m} $$

where $p_{(123)}$ is a $3 \times 3$ matrix representing the first three columns of the projection matrix $P$, and $p_{(4)}$ is the fourth column.

The observation model describes the process of implicitly representing observed features $A$ in a 3-D parameterization as a function of the systems state $s$ parameters by inferring constraints to the environment structure and consequently to the features 3-D location and motion.
the state transition function $f$ is applied to $s_k$, retrieving $s_{k+1} = f(s_k, n_k)$. $\Lambda_k$ is the input set of features in pixel coordinates acquired at the instant $k$, whereas $\hat{\Lambda}_{k+1}$ is the set of features in following instant $k + 1$ and the observation model output. The observation model can thus be written as a function $h$:

$$\hat{\Lambda}_{k+1} = h_{k+1}(s_{k+1}, \Lambda_k) \quad (9)$$

Consider that the set of features $\Lambda_k$, 3-D locations on the TSS surface, are the intersections of rays leaving the camera centre $r_{WC}^{k}$ with the cylinder wall, piercing the 2-D image plane where all features lie. With the information present in the state vector $s_k$, i.e. camera pose $x_k$ and cylinders section parameters $y_k$, one can estimate the 3-D location of every feature in set $\Lambda_k$ through a back-projection methodology, and compute the re-projection of this 3-D coordinates with the next instant $k + 1$ state vector $s_{k+1}$ to a 2-D image plane, thus acquiring $\hat{\Lambda}_{k+1}$ features.

3 Dewarping

The process of dewarping the interior of the tube is done by finding the transformation between the pixels in the “open” images and the pixels in the “closed” scenes viewed by the camera inside the TSS. Figure 4 shows that relation. Knowing that relation, each pixel in the “open” image has a corresponding 3D position in reference to the camera’s centre. The projection of that 3D point in the “closed” image is the respective pixel in the “open” image. A final mosaic is stitched showing the full inside texture of the tube.

4 Experimental Results

Two experiments have been conducted to test the proposed methodology. In the first experiment the tubular shape is simulated in VRML. The VRML based system imply using image processing, namely features (e.g. SIFT or SURF) detection and matching, enhanced by the RANSAC methodology for outliers removal. Figure 1 shows various steps and results of the process. The filtered nature of the estimated tubular structure and camera motion implies a dewarping which has small variations, expansion or compression, along the tube length. In Figure 1(d) the small variations along tube length are compensated vertically (the small black regions in the mosaic indicate this compensation).

The second experiment consists of evaluating the developed algorithm with a set of images retrieved from a real camera navigation inside a textured TSS. The setup and the results can be seen in Figure 2. The resulting mosaic, Figure 2(d), shows some horizontal oscillation denoting under- or over-estimation of camera rotation. This is due to biased observations such as the ones associated to an unbalanced number of features around the camera.

5 Conclusion

This paper aimed at presenting a robust model to solve the Simultaneous Localization and Mapping problem with a priori structural environment knowledge, assumed to be a TSS. The cameras pose estimate and the camera trajectory can be recovered, up to a scale factor, assuming the constant speed dynamic model, allowing a posterior description of the environment structure and its reconstruction.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE / EEI / LA0009 / 2013, by the FCT project PTDC / EEACRO / 105413 / 2008 DCCAL, and by the FCT project EXPL / EEI-AUT / 1560 / 2013 ACDC.

References

On Compression-Based Text Authorship Attribution

David Pereira Coutinho
davidpc@cc.isel.pt

Mário A. T. Figueiredo
mario.figueiredo@lx.it.pt

Instituto Superior de Engenharia de Lisboa and Instituto de Telecomunicações
Lisboa, Portugal

Instituto Superior Técnico and Instituto de Telecomunicações
Lisboa, Portugal

Abstract

Choosing an appropriate set of features that allows a machine learning algorithm to accurately solve a given problem is arguably one of the most difficult tasks in text classification. Most state-of-the-art approaches involve careful feature engineering following a preprocessing stage, which may be too expensive in the emerging context of massive collections of electronic texts. In this paper, we propose efficient methods for text classification, based on information-theoretic dissimilarity measures, which are used to define dissimilarity-based representations. These methods dispense with any feature design or engineering, by mapping texts into a feature space using universal dissimilarity measures; in this space, classical classifiers (e.g. nearest neighbor or support vector machines) can then be used. The reported experimental evaluation of the proposed methods, on a benchmark authorship attribution problem, reveals that it outperforms previous methods, despite being much simpler, in the sense that it does not require any pre-processing or feature engineering.

1 Introduction

Text classification (or categorization) is the problem of assigning a text to one or more of a predefined set of classes. Examples of applications are: (i) topic classification, where the task is to decide which topic(s) is (are) addressed in a text; (ii) sentiment analysis, which is the task of automatically classifying a text, not in terms of topic, but according to the overall sentiment it expresses; (iii) authorship attribution (AA), where the task is to assign a text of an unknown author to one of a set of possible authors. Classical techniques for these (and other) text classification tasks are based on statistical and computational tools that require careful feature engineering and sophisticated preprocessing, which may become prohibitive in the emerging context of massive collections of electronic texts, such as product reviews and e-mail messages. Defining a similarity measure between texts (or, more generally, finite sequences of symbols) that allows addressing classification problems, without explicitly modeling their statistical behavior, is a fundamental problem in this context. In this paper, we aim at dispensing with the need for any feature design or engineering. For that purpose, we partially follow our previous work [5], in that we use compression/parsing-based feature extractors that don’t require any human intervention or feature design. We adopt a dissimilarity space approach [4], in which each text is characterized by the vector of its (dis)similarities with respect to the other texts in the training set. Finally, standard nearest-neighbor (NN) and support vector machines (SVM) are used as classifiers. Our experimental results reveal that our approach outperforms previous methods.

2 Compression-based dissimilarity measures

In recent years, much work has been done concerning the design and development of information-theoretic dissimilarity measures. Most of the proposed approaches were developed in a Kolmogorov complexity framework (e.g. [3]), but an alternative approach can be based on Shannon’s relative entropy [6].

2.1 The Normalized Compression Distance

One of the best known compression-based dissimilarity measures for text is the normalized compression distance (NCD), proposed by Li et al. [3]. NCD approximates the (non-computable) Kolmogorov complexity of a string \( x \) by the length of a compressed version of \( x \), using off-the-shelf compression algorithms such as gzip or bzip2, and it is defined for any pair of strings \( x \) and \( y \) as

\[
NCD(x, y) = \frac{C(x \cup y) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}},
\]

where \( C(x) \) is the length of string \( x \) after being compressed by a lossless compression algorithm, and \( x \cup y \) denotes the concatenation of strings \( x \) and \( y \). The authors demonstrate that it is a metric and claim that it minimizes every computable distance in a certain class. NCD ranges from 0 to 1 + \( \epsilon \), where 0 corresponds to \( x \) and \( y \) being identical, and 1 means maximum dissimilarity. The constant \( \epsilon \) is an upper bound due to imperfections in the compression algorithms, but is unlikely to be above 0.1 for most standard compressors [3].

2.2 The Ziv-Merhav Relative Entropy Estimate

A method for estimating the relative (Shannon) entropy between pairs of sequences of symbols was introduced by Ziv and Merhav (ZM) [6] and has been used as a dissimilarity measure for universal classification [5]. The ZM method is based on the incremental Lempel-Ziv (LZ) parsing algorithm and on a variation thereof, known as cross-parsing. Combining these two algorithms, Ziv and Merhav proposed an estimator of the relative entropy between two ergodic sources producing the sequences \( x \) and \( y \), which can be used as a dissimilarity measure between those sequences. Specifically, they proved that for two finite order (of any order) Markovian sequences of length \( n \), the quantity

\[
\Delta(z|x) = \frac{1}{n} \left[ c(z|x) \log_2 n - c(z) \log_2 C(z) \right]
\]

converges, as \( n \to \infty \), to the relative entropy between the two sources that emitted the two sequences, where \( c(z) \) denotes the number of phrases resulting from the self-parsing of \( z \) and \( c(z|x) \) is the number of phrases resulting from cross-parsing \( z \) with respect to \( x \). Roughly speaking, we can interpret \( (1/n) c(z) \log_2 C(z) \) as a measure of complexity of the sequence \( z \), obtained by self-parsing, thus providing an estimate of its entropy, while \( (1/n) c(z|x) \log_2 n \) can be seen as an estimate of the code-length obtained when coding \( z \) using a model for \( x \). The difference between the two quantities thus provides a measure of how different the distributions that produced the two sequences are.

2.3 The Cross-Parsing Distance

The use of the Ziv-Merhav relative entropy estimate is not directly applicable in some scenarios, namely because it is defined for sequences of the same length \( n \). When generalizing this definition to sequences of different lengths, several problems arise, with the size of the “model” sequence \( x \) having a significant impact [2]. To overcome this difficulty, Helmer et al. [2] recently introduced the cross-parsing distance (CPD), which is a semi-metric (i.e., of all the conditions that have to be satisfied by a metric, it only does not satisfy the triangle inequality) defined for any pair of strings \( x \) and \( y \) (of length respectively \( |x| \) and \( |y| \)), as

\[
dist_{CPD}(x, y) = \frac{1}{2} \left( \frac{|s(x|y) \setminus x|}{|x|} + \frac{|s(y|x) \setminus y|}{|y|} \right),
\]

where \( s(x|y) \) denotes the multiset of all phrases resulting from the cross-parsing of \( x \) with respect to \( y \) and \( s(y|x) \) \( \setminus \) \( y \) denotes the removal of a single instance of \( y \) from the multiset \( s(x|y) \) (if one exists, even if multiple copies exist). If the first not yet parsed symbol in \( x \) is not found in \( y \), then the parsing is simply the symbol itself (see Helmer et al. [2] for details).

We will use in this paper as a variant of CPD, a modified version of \( \text{dist}_{CPD} \) which we call \( CP_{dist} \), defined as

\[
CP_{dist}(x, y) = \frac{1}{2} \left( c(x|y) - 1_{x=y} \right) \frac{|x|}{|x|} + c(y|x) - 1_{x=y} \frac{|y|}{|y|},
\]

where \( 1_{x=y} \) is the indicator function that is 1 if \( x \) and \( y \) are identical, and 0 otherwise.
2.4 Incremental Cross-Parsing Algorithm

An implementation of the cross-parsing algorithm based on the LZ77 (here termed CP77) was proposed by Pereira Coutinho and Figueiredo [5]; this implementation uses a 2 Mbyte dictionary and a 256 byte look ahead buffer (LAB), where the dictionary is static and only the LAB slides over the input sequence (for details see [5]). However LZ77 is itself an incremental parsing algorithm; thus, following the same idea, we propose a new implementation, based on our first implementation but using incremental dictionary updates (termed CP77inc). Now, the cross-parsing of string \( z \) with respect to the string \( x \) involves some details, which we briefly describe; it uses one sliding window to hold both the dictionary \( D_x \) and the LAB \( lab_x \) of the model string \( x \). In addition, it uses another (smaller) sliding window to hold the LAB \( lab_y \), for the unknown string \( y \). Notice that the dictionary \( D_x \) is empty at the beginning. Then, a loop is repeated until the end of \( z \) is reached: the cross-parsing of \( z \) given \( x \); the self parsing of \( x \) including dictionary update as long as \( x \) lasts. This makes sequences of different lengths allowed, by stopping the dictionary update whenever the end of \( x \) is reached and keep using it as a “static” dictionary. Every time the loop is executed, a counter \( c_{xy} \) is incremented. Finally, we call the method of relative entropy estimate via definition (2) as ZMMinc, which uses the proposed algorithm CP77inc.

3 Proposed dissimilarity-based classification

At the core of dissimilarity-based methods for classification is the computation of pairwise dissimilarities between the object (e.g. text) to be classified and a set of (or all) objects (e.g. texts) in the training set. Of course, there are several ways to use dissimilarity values to define a classifier, the simplest of which is arguably to use a \( k \)-NN classifier; in this case, the object to be classified is simply assigned to the majority class in its \( k \) nearest (in the adopted similarity measure) neighbors (with some rule to break ties). A more sophisticated approach is offered by the dissimilarity space approach [4], which uses the dissimilarity values as features that characterize the object to be classified, based on which several different types of classifiers can be used, namely \( k \)-NN in the dissimilarity space or support vector machines (SVM).

Let us consider a training collection of objects (texts) \( X = \{x_1, ..., x_n\} \), where each object belongs to some set \( X \) (e.g. the set of finite length strings of some finite alphabet \( Z \)), and some dissimilarity measure between pairs of objects, \( D : X \times X \to \mathbb{R} \). In the dissimilarity-based approach, each object (either in the training set or a new object to be classified after training) is represented by the vector of its dissimilarities with respect to the elements of \( X \) (or a subset thereof). That is, the training set in the so-called dissimilarity space becomes

\[
D = \{d_1, ..., d_n\},
\]

where

\[
d_i = \begin{bmatrix}
D(x_i, x_1) \\
\vdots \\
D(x_i, x_n)
\end{bmatrix} \in \mathbb{R}^n.
\]

In this paper, we propose to use a dissimilarity space approach, where the representations are built by using the dissimilarity/distance measures described in the previous section (see Figure 1). Once in possession of a dissimilarity-based representation of a training set, any standard classification method can be used.

For example, when using a \( k \)-NN classifier, given a new object, its dissimilarity vector \( d_k \) is built and distances in the dissimilarity space are computed (Euclidean distances, for example) between the new vector \( d_k \) and all the vectors in \( D \). Then, the new object is classified in the most common class amongst its \( k \) nearest neighbors.

An important aspect of dissimilarity-based approaches is that very few conditions are put of the dissimilarity measure; namely, it doesn’t have to be metric, it doesn’t even need to be symmetric [4].

4 Experiments

Experiments were carried out using both \( k \)-NN and (linear) SVM classifiers on the dissimilarity space, with the accuracy assessed by leave-one-out cross-validation (LOO-CV), and using the English Corpus recently introduced by Ebrahimpour et al. [1]. This corpus\(^1\) contains 168 short stories with undisputed authorship by seven English writers of the late 19th century and early 20th century; each story is truncate to approximately the first 5,000 words. Table 1 shows the accuracy results for that corpus when using NCD, CPdist and ZMMinc as dissimilarity measures. Our best method (ZMMinc and SVM with optimized \( C \) parameter), only misclassifies 2 out of the 168 texts.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Baseline [1]</th>
<th>NCD</th>
<th>CPdist</th>
<th>ZMMinc</th>
<th>NCD</th>
<th>CPdist</th>
<th>ZMMinc</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>96.4</td>
<td>94.5</td>
<td>91.3</td>
<td>87.5</td>
<td>99.2</td>
<td>95.2</td>
<td>98.8</td>
</tr>
</tbody>
</table>

5 Conclusions

Achieving good accuracy in text classification usually requires careful feature engineering and complex preprocessing stages, which may become prohibitive in the emerging context of classification massive sets of electronic texts. In this paper, we proposed methods for automatic text classification using information-theoretic dissimilarity measures, based on universal data compression algorithms, which bypass the feature design and preprocessing stages.

On an authorship attribution problem, experiments were done using several dissimilarity measures and the best of the proposed methods outperformed the state-of-the-art approaches.

References


\(^1\)Available at http://promo.net/pg
Region clustering using colour tuned keypoints

Miguel Farrajota
mfarrajota@ualg.pt
J.M.F. Rodrigues
jrodrig@ualg.pt
J.M.H. du Buf
dubuf@ualg.pt

Vision Laboratory, Inst. for Systems and Robotics,
LARSyS, University of the Algarve
Campus de Gambelas, 8005-139 Faro, Portugal

Abstract

Coloured regions can be segregated by using colour-opponent mechanism, colour contrast, saturation or luminance. This paper focuses on clustering coloured regions by using colour tuned end-stopped cells. Colour information is coded in separate colour channels to convey differently coloured regions. Then, by using multi-scale cortical end-stopped cells tuned to colour, this information is coded in all channels by mapping it to the multi-scale peaks. Finally, unsupervised clustering is achieved by analysing the branches of these peaks and by linking them together on the basis of their colour, saturation and luminance information.

1 Introduction

Distinguishing visual patches in a scene is important for trying to achieve better and faster object detection and recognition results in computer vision [1]. The development of perceptual grouping algorithms is motivated by the idea that a correct organisation of the visual scene into meaningful regions and/or feature groups will lead to more reliable and efficient methods for object detection, tracking and scene interpretation. These methods are related to two topics of computer vision, namely perceptual organisation and image segmentation. Perceptual organisation deals with the multi-scale peaks. Finally, unsupervised clustering is achieved by tuning to colour, this information is coded in all channels by mapping it to the multi-scale peaks. Finally, unsupervised clustering is achieved by analysing the branches of these peaks and by linking them together on the basis of their colour, saturation and luminance information.

2 Colour coding cells and keypoint detection

The method applies a colour gain function in conjunction with a high-pass Butterworth function to the hue channels and a low-pass Butterworth function to the chroma channel for shades, in order to obtain colour- and shade-tuned channels. Here we use the CIE L*a*b*  colour space: let image \( f(x,y) \), with \( N \times M \) pixels, be defined as \( (L^*, C^*, H^*) \). We divide the hue circle into \( N_h = 8 \) equal ranges (channels). The hue gain functions are Gaussians tuned to specific hues \( \phi_j \); \( \phi_j = j \times 360/N_h \) and \( j = \{1, 2, 3, \ldots, N_h\} \). Each hue \( H^*(x,y) \) in each channel \( \phi_j \) is weighted by a Gaussian gain function \( G_{\phi_j} (x,y) = \exp(-((H^*_j(x,y) - \phi_j)^2/(2\sigma^2))) \), with \( \sigma = 360/N_h \). To the chroma \( C^* \) component it is applied a low-pass Butterworth function \( BW(x,y) = \sqrt{1 + (1 + C^*_j(x,y)/K)^2} \) is applied. A high-pass \( BW \) function is applied to \( G_{\phi_j} \), which yields the colour contrast \( \Psi_{CC_j}(x,y) = G_{\phi_j}(x,y) \times (1 - BW(x,y)) \), with \( CC_j = \{1, \ldots, N_h\} \), \( \eta = 3 \) and \( K = 6 \). In order to weaken colours with low saturation, a low-pass

Finally, the luminance is not processed: \( \Psi_L = L^*(x,y) \). All colour and shade responses \( \Psi \) are normalised between 0 and 1.

The basic principle of multi-scale colour coding is based on Gabor quadrature filters which provide a model of cortical simple cells [7]. In the spatial domain \( (x,y) \) they consist of a real cosine and an imaginary sine, both with a Gaussian envelope. Responses of even and odd simple cells, which correspond to real and imaginary parts of a Gabor filter, are obtained by convolving the input image with the filter kernel, and are denoted by \( R_{j,k}^0(x,y) \) and \( R_{j,k}^1(x,y) \), being \( s = \{8, 12, 16, \ldots, 2 \min\{N, M\}\} \) the scales with half-octave increments, the orientations \( i = \{0, N_d - 1\} \) with \( N_d \) the number of orientations (here 8), and \( h = \{CC_j, SC, L\} \) the colour \( (CC_j) \), saturation \( (SC) \) and luminance channel \( (L) \). Responses of complex cells are modelled by the modulus \( C_{i,j,k}(x,y) \) [7] and normalised between 0 and 1.

There are two types of end-stopped cells: single and double [7]. These cells are combined, using \( C_{i,j,k} \) in order to obtain the cells’ responses in all colour channels. If \( i > 5 \) denotes the suppression of negative values, and \( C_i = \cos(\theta_i/\pi/N_d) \) and \( S_i = \sin(\theta_i) \), then single end-stopped cells are modelled by

\[
S_{i,j,k}(x,y) = \left[ C_{i,j,k}(x+{dS_{i,j,i,y} - dC_{i,j}}) - C_{i,j,k}(x-{dS_{i,j,i,y} + dC_{i,j}}) \right]^{+},
\]

and double end-stopped ones by

\[
D_{i,j,k}(x,y) = \left[ C_{i,j,k}(x) \times C_{i,j,k}(y) - C_{i,j,k}(x) + C_{i,j,k}(x) \right]^{+},
\]

with \( CS_{i,j,k} = \frac{1}{2} C_{i,j,k}(x+{2dS_{i,j,i,y} - 2dC_{i,j}}) + \frac{1}{2} C_{i,j,k}(x-{2dS_{i,j,i,y} + 2dC_{i,j}}) \). Distance \( d \) is scaled linearly with the filter scale \( s = d/5 \).

Hubel [4] reported some end-stopped cells which did not respond at all to long lines, and he coined them as completely end-stopped cells. Although double end-stopped cells convey information concerning certain patterns, completely end-stopped cells also convey information if the stimulus area is larger than the activation region of the receptive field (RF). Based on such cells, we apply the same property for detecting circular regions. If \( CS_{i,j,k}(x) > 0.55 \times C_{i,j,k}(x) \), then the response of the completely end-stopped cell \( CD_{i,j,k} \) is inhibited: \( CD_{i,j,k}(x) = 0 \). If the cell’s response is not inhibited, it assumes the same value as the double end-stopped cell: \( CD_{i,j,k}(x) = D_{i,j,k}(x) \).

In this scale space of end-stopped colour cells we look for peaks ("extrema") at each scale which can code differently coloured regions. Cell responses are summed over all orientations: if \( A = \{S, D, C\} \), then \( \Lambda_{i,j,h} = \sum_{i=0}^{N_d-1} \Lambda_{i,j,h}/N_d \); a threshold \( T_i = 0.2 \) is applied to inhibit small responses. The maximum responses of all \( h \) channels are combined, i.e., \( \Lambda_h = \max_h \{\Lambda_{i,h}\} \), and the local extrema are detected: \( E^1_{\Lambda} = \text{peak} \{\Lambda_{h}\} \) are the peaks of the local maxima of each detected region. Each region \( \Lambda_h \) is now assigned a label that corresponds to the \( \phi_j \) of the \( max_h \), plus two labels for saturation and luminance contrast. The result is an image that has \( N_{\phi} + 2 \) labels: \( \Gamma(x,y) \) is composed by the maximum colour responses of channels \( h \) of \( \Lambda_h \). This image classifies the keypoints (extrema) with respect to colour.

3 Keypoint clustering

The clustering process consists of four steps: (1) trees of keypoints are clustered in all colour channels in a top-down way; (2) trees which mainly consist of keypoints without (inhibited) completely end-stopped responses are separated from those with such responses – because inhibited responses are due to very elongated regions whereas the other ones are due to circular or semi-circular regions; (3) trees with inhibited completely end-stopped responses within the same hue range \( \phi \) are clustered.
basis of saturation, luminance and spatial continuity; and (4) the resulting clusters are linked to other clusters belonging to neighbouring hue ranges.

Keypoints of the same hue range $\phi$ and from all scales are combined into trees. We apply a multi-scale tree structure in which one keypoint at a coarse scale is related to one or more keypoints at a finer scale, which can be slightly displaced. This relation is modelled by down-projection using grouping cells with a circular axonic field, the size of which ($\lambda$) defines the region of influence; see [2]. Resulting trees mainly composed of responses of completely end-stopped cells are then separated from those with inhibited responses. If a tree only comprises keypoints with “completely” responses, then it is considered as a final cluster and it is excluded from further processing.

In the following step, trees of the same colour are clustered with respect to saturation, luminance and spatial continuity. Let $T_{A\phi}$ and $T_{B\phi}$ be two trees with the same colour. For each tree, the means $C^*_{A\phi}$ is the number of scales a particular tree is composed of. If $C^*_{A\phi} > 0.2$ and $C^*_{A\phi} > 0.25$, then the spatial continuity is checked. A binary map $B_{j}$ is derived from the colour maps $CC_{j}:

$$B_{j}(x,y) = \begin{cases} 1 & \text{if } CC_{j}(x,y) \geq 0.7 \\ 0 & \text{if } CC_{j}(x,y) < 0.7. \end{cases} \quad (3)$$

Now, between two keypoints of each tree, on a straight line connecting both extrema, if $B_{j}(x,y) = 0$ at 6 or more consecutive pixels, the link is considered invalid until a valid one from all pairs is detected, and the two trees are grouped together. This process is repeated between all trees and clusters of trees until all possible links have been checked.

Finally, clusters of all colours are evaluated and combined in cases where the hue of the underlying patch lies between two hue ranges. The $\Psi_{3C}$ and $\Psi_{L}$ channels are used in this step. In case of two clusters with neighbouring colours $\phi_{l}$ and $\phi_{r}$, with $|l - r| \leq 1$, any two trees from each cluster are compared as in the previous clustering step, where the saturation and luminance means are computed and validated. If positively validated, the minimum distance between the closest two keypoints from both trees is calculated, and if this distance is less than 13 pixels, both clusters are merged. In the case of clusters already composed by two different colour ranges (\(\phi_{l}, \phi_{r}\)), only clusters within the same hue range are considered for validation and merging. As before, this process is repeated until all links have been checked.

## 4 Discussion and results

This paper presented a biologically inspired method for keypoint detection and clustering. Keypoints labelled by colour provide stable features across images and scales. Also, by exploiting the smoothness of the Gaussian scale space, clustering keypoints into trees constitutes a fairly simple and robust solution for basic grouping of features, with encouraging stability over scales. Likewise, better clustering results can be achieved by using information such as colour continuity, saturation and luminance. More complete and distinct sets of features for coloured regions can be obtained. The usefulness of such a clustering is certainly more focused on object comparison and recognition [5], because groups of features convey better spatial information and this can result in improved matching and results of the method demonstrate the applicability and usefulness of keypoint grouping. Figure 1 shows results for two images. In the second column, the maximum colour responses $\Gamma$ show a clear colour sampling and clustering of pixels on the basis of hue. The third column shows all keypoints at all scales, with the same colour coding as in the previous column. Examples of the overall clustering process are illustrated in the right two columns. Shown are the clustered keypoints of the surfer’s wetsuit and the big wave to the left. Also shown are those of the frog and yellow flower, both consisting of two clusters because of two colours.

It can be seen that the results of both images are good but not yet perfect. Although keypoints of the surfer and the greenish wave are correctly clustered, the frog and flower are more difficult. About five keypoints around the frog are not correct. The line of keypoints at the tip of the very thin leaf in the top-right corner (in the fourth column) is missing, but this is likely caused by the different colour, i.e., the very grayish yellow/orange. In general, it is difficult to cluster colours with low saturation levels. Nevertheless, although some details are not perfect, overall results are good enough to be used in matching and segmentation processes.

## Acknowledgments

Portuguese Foundation for Science and Technology (FCT) project LARSyS (PEst-OE/EEI/LA0009/2013), project Blavigator RIPD/ADA/109600/2009 and PhD grant to author MF (SFRH/BD/79812/2011).

## References


AAM Based Vocal Tract Segmentation from Real-Time MRI Image Sequences

Samuel Silva
sweet.ua.pt/sss
António Teixeira
sweet.ua.pt/ajst

Institute of Electronics Engineering and Telematics of Aveiro (IEETA)
Department of Electronics, Telecommunications and Informatics (DETI)
University of Aveiro
Aveiro, PT

Abstract

The study of the dynamic aspects of speech production has been performed, in recent years, using Real-Time Magnetic Resonance Imaging (RT-MRI). One of the main challenges concerns how to deal with the large amount of images resulting from these studies and extract relevant data for analysis such as the vocal tract contours.

Based on active appearance models (AAMs), a method is proposed to segment the vocal tract from midsagittal RT-MRI sequences. First evaluation results against manual segmentations performed by two observers show very good accuracy results.

1 Introduction

Real-time MRI has a wide range of applications such as the analysis of swallowing disorders [6] or the study of the dynamic aspects of speech [11].

Analysis of the vocal tract configuration has traditionally been performed at pixel level [3] or considering specific articulators (e.g., tongue dorsum [4, 7, 12]). Segmentation of the full vocal tract has already been presented by different authors, but considering raw image acquisition data [1], restricted application to single speaker data [5, 14] or grid based analysis, defining search paths, over the image, for which local intensity minima are computed and used to detect vocal tract limits [9].

To address the task of segmenting a large database of RT-MRI sequences (containing 12000+ images), acquired to study EP vowels, a segmentation method is proposed, based on active appearance models (AAMs). This method solves problems detected on a previously proposed method [10] by providing unsupervised operation and clear separation between articulators, even in case of strong contact (e.g., tongue and hard palate). The proposed method requires a single initialization step per speaker and can work unsupervised thereafter.

This paper is organized as follows: section 2 describes the main aspects of training and applying the proposed segmentation method; section 3 presents data regarding accuracy evaluation against manual segmentations performed by observers; finally, section 4 presents some conclusions and ideas for further work.

2 Methods

The following sections provide a brief description of the methods used for image acquisition, AAM training and vocal tract segmentation.

2.1 Image Acquisition

Using RT-MRI [13], at a frame rate of 14 images/s, the corpus, acquired for three speakers (CM, CO and SV), includes: 1) the five EP nasal vowels (e.g. [p̃a], [p̃am], [p̃a], [p̃am], [p̃a]) in word initial, medial and final positions (e.g. “ampa, pampa, pam” [p̃ep], [p̃ep], [p̃ep]); and 2) the eight oral vowels (e.g. [p̃a], [p̃a], [p̃e], [p̃i], [p̃o], [p̃u]) inserted in CV1CV2 sequences (e.g. [p̃a][p̃a], [p̃a][p̃a]). Synchronized audio was recorded during image acquisition and annotated manually using Praat (http://www.fon.hum.uva.nl/praat/).

2.2 Segmentation

In simple terms, active appearance models (AAMs) rely on a point description model (PDM), describing the geometric properties of the object to segment and an associated texture model, describing the appearance expected characteristics of the object to segment. The mean geometric model is then positioned over the object to segment, in new images, and it is iteratively adjusted to attain the best possible compromise considering the shape and appearance modelled. Given the lack of space, for a detailed description the reader is forwarded to [2].

The image databases for articulatory studies do not usually include many speakers, given the high cost of each acquisition session. Instead, they include a large amount of data per speaker. Therefore, our purpose was to propose a segmentation method that, based on a small number of annotated images, could perform unsupervised processing of all the data.

For model training, 30 images were chosen, covering the three speakers and the different sounds present in the database. These do not cover all the possible configurations adopted by the vocal tract, during running speech, but cover the most important. The training images were manually annotated defining a set of twenty-six landmarks in notable positions of the vocal tract (fig. 1): three for the lower lip, two for the lingual frenulum, one for the tongue tip, seven for the tongue, four for the pharynx/larynx, three for the velum, three for the hard palate and three for the upper lip.

For each speaker, the mean model initial position was manually defined for one of the images and the final segmentation inspected to ensure if it was correct. Since there is spatial coherence among all image sequences for each speaker, this initial position was used for the remaining images allowing unsupervised segmentation.

Figure 2 presents some segmentation examples for images of different sequences and speakers. As can be observed, the segmentations correctly adjust to the vocal tract, resulting in smooth contours.

3 Evaluation

To assess the accuracy of the proposed segmentation method, a set of 50 images was chosen from the database (not including any of the training images) and manually segmented by two observers, a linguist (OBS1) and a radiographer (OBS2). Some examples of observer vs automatic segmentations are presented in figure 3.

The observer segmentations were compared with those obtained with the proposed method using the Dice similarity coefficient[8]:

\[
DSC = \frac{2 |A \cap B|}{|A| + |B|} \tag{1}
\]

where A and B are the two compared regions. Table 1 shows the mean and standard deviation of the DSC values obtained for each observer.

The mean DSC values show evidence that the proposed method exhibits good accuracy.
Figure 2: Examples of segmentations obtained using the proposed method.

<table>
<thead>
<tr>
<th></th>
<th>DSC (%)</th>
<th>µ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBS1</td>
<td>83</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OBS2</td>
<td>80</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Comparison of proposed segmentation method with observer segmentations: mean (µ) and standard deviation (σ) values for DSC.

4 Conclusions

An AAM based segmentation method is proposed to perform vocal tract segmentation from RT-MRI image sequences. Evaluation against manual segmentations performed by observers yielded good accuracy results. Further evaluation can still be performed regarding precision, regional accuracy and performance.

Given the large amount of data involved (image sequences, segmented contours, audio signal, sound annotations, etc.), its exploration and usage (including third-parties, e.g. in hospital environments) would greatly profit if a repository was available, allowing data storage, remote data processing and data retrieval. To move towards that goal, a Cloud based approach is being considered.

5 Acknowledgements

Research partially funded by FEDER through the Program COMPETE and by National Funds (FCT) in the context of HERON II (PTDC/EEA-PLP/098298/2008), IEETA Research Unit funding FCOMP-01-0124-FEDER-022682 (FCT-PEst-O/EEI/UI0127/2011) and project Cloud Thinking (funded by the QREN Mais Centro program, ref. CENTRO-07-ST24-FEDER-002031).

References


AFM based-force spectroscopy as a functional diagnostic nanotool for hematological diseases

Filomena A. Carvalho¹
filomenacarvalho@fm.ul.pt

Alice Tavares, Mafalda Teodoro²
alice.tavares@chln.min-saude.pt

Gabriel Miltenberger-Miltenyi¹
gmittenyi@fm.ul.pt

Nuno C. Santos¹
nsantos@fm.ul.pt

¹Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal
²Servicio de Imuno-Hemoterapia, Hospital de Santa Maria, Centro Hospitalar Lisboa Norte, Lisboa, Portugal

Abstract

Glanzmann thrombasthenia (GT) is a rare hereditary hematological disease associated to the membrane glycoprotein αIIbβ3, the integrin receptor for fibrinogen in platelets. Using atomic force microscopy (AFM)-based force spectroscopy, we have previously characterized at the single-molecule level and compared the binding of fibrinogen to human platelets and erythrocytes [1]. The aim of this study was to set AFM based-force spectroscopy as a functional diagnostic tool for hematological diseases. The interaction between fibrinogen and platelets from GT patients was studied using fibrinogen-functionalized AFM tips. Our results show that, if the patient has a mutation on the β3 subunit-associated gene, significant reductions both on the frequency of fibrinogen-platelet binding and on its force occur. When the GT patient has a homozygous mutation on the αIIb subunit gene, a significant reduction of the frequency of the fibrinogen-platelet interaction arises, but not on its strength. The relation of these results with the clinical data demonstrates the applicability of AFM as a highly sensitive nanotool for the functional evaluation of the outcome of genetic mutations resulting in hematological diseases.

1 Introduction

Fibrinogen plays a central role in the mechanisms of coagulation and thrombosis. It is a high molecular weight plasma adhesion protein, a glycoprotein dimer composed of three pairs of non-identical polypeptide chains (α, β and γ), and a biomarker of inflammation. It is constitutively expressed exclusively in hepatocytes and is inducible by interleukin-6 (IL-6) as part of the acute phase reaction [2]. Fibrinogen has a plasma half-life of 3-5 days. In the final step of the coagulation cascade, it is converted into fibrin under the action of thrombin. Its binding to the αIIbβ3 receptor in activated platelets is a key step for platelet aggregation. There is a growing number of reports evidencing that increased fibrinogen concentration is a significant risk factor for several cardiovascular and cerebrovascular disorders [3,4]. Increased levels of fibrinogen result in changes in blood rheological properties, such as increases in plasma viscosity, erythrocyte aggregation and platelet thrombogenesis, along with alterations in vascular reactivity and compromised endothelial layer integrity [5]. These alterations exacerbate the complications in peripheral blood circulation during cardiovascular conditions such as hypertension, diabetes and stroke. In addition to affecting blood viscosity by altering plasma viscosity and erythrocyte aggregation, growing experimental evidences indicate that fibrinogen alters vascular reactivity and impairs endothelial cell layer integrity by binding to its endothelial cell membrane receptors and activating signaling mechanisms [5,6].

Erythrocyte aggregation has become an issue of increasing interest due to its pathogenic implications in thrombus formation, both at venous and arterial level. This parameter is used especially to assess the risk of primary or secondary cardiovascular events [4], since it is influenced mostly by fibrinogen and plasmatic lipids. Hadengue showed that in hypertension and hypercholesterolemia, the increase in erythrocyte aggregation could be attributed to an increase in the concentration of plasma fibrinogen [7]. In clinical practice, erythrocyte aggregation is basically used to assess thrombotic tendency in subjects with already known cardiovascular risk factors, such as obesity, dyslipidemia, smoking habit, hypertension or diabetes mellitus, as in patients who had a primary cardiovascular episode and the risk of suffering a secondary event needs to be evaluated. It is in such clinical situations and in a non-acute phase of the disease that it is more useful to determine erythrocyte aggregation as an indicator of cardiovascular risk [4].

Biological systems can only be fully understood if their structure and mechanism are known. This underscores the importance of structural biology, the science investigating the structure and function of the components of living systems. A thorough study of the interactions between erythrocytes and fibrinogen molecules to understand the observed increase of erythrocyte aggregation on patients with some cardiovascular risk factors is therefore of paramount importance.

Until the recent publication of an article by us [1], the prevailing hypothesis for the mechanism of fibrinogen-induced erythrocyte hyperaggregation was that it could be caused by a nonspecific binding mechanism [6], despite the published data on the changes in erythrocyte aggregation during hypertension pointed to the existence of other mechanism(s) [8].

The use of nanotechnologies for medical applications raises high expectations regarding diagnosis, drug delivery, gene therapy and tissue engineering. There is an increasing number of reports using atomic force microscopy as a nanodiagnostics tool for patient cells. Beside its direct relevance on the identification of the fibrinogen receptor on erythrocytes and of a pharmacological strategy to inhibit it, our recent work was also a demonstration of the applicability and validation of the AFM-based force spectroscopy technique as a highly sensitive and low operation cost nanotool for the diagnostic and unbiased functional evaluation of the severity of hematological diseases arising from genetic mutations [1]. In that work, based on force spectroscopy measurements using an atomic force microscope, we reported the existence of a single-molecule interaction between fibrinogen and an unknown receptor on the erythrocyte membrane, with a lower but comparable affinity relative to platelet binding. The receptor identified by us in erythrocytes is not as strongly influenced by calcium and epifibatide (an αIIbβ3 specific inhibitor) as the platelet receptor. However, its inhibition by epifibatide indicates that it is an αIIbβ3-related integrin. The results obtained for a Glanzmann thrombasthenia (a rare hereditary bleeding disease caused by αIIbβ3 deficiency) patient showed (for the first time) an impaired fibrinogen-erythrocyte binding. This methodology enables a unique unbiased functional nanodiagnostic (at the single-molecule level) of the severity of the disease for Glanzmann thrombasthenia patients. Correlation with genetic sequencing data demonstrated that one of the units of the fibrinogen receptor on erythrocytes is a product of the expression of the β3 gene, found to be mutated on this patient [1].

More recently, we used the same AFM-based methodology, complemented by fluorescence spectroscopy and zeta-potential measurements, to assess the changes on fibrinogen-erythrocyte interaction upon in vivo erythrocyte ageing [10]. Our data indicate that younger erythrocytes are the main responsible for some cardiovascular diseases associated with an increase on the fibrinogen content in blood.

Figure 1: (A) Scheme of an atomic force microscope [9]. (B) Picture of the atomic force microscope coupled with a fluorescent inverted microscope used on this experiment (JPJ Instruments, Berlin).
RecPad 2013  Page 102 of 147

2 Methods

2.1 AFM-based force spectroscopy

Understanding the forces between a single biomolecule and its receptor is a challenging task. This can be achieved by force spectroscopy, using an atomic force microscope [11,12]. Force spectroscopy allows the study of the mechanical properties of single polymer molecules or proteins, or individual chemical bonds. It is performed by pulling on the system under scrutiny with controlled forces. As a single-molecule technique, as opposed to typical ensemble techniques, it allows the determination of specific molecular properties, even for interactions with a short life-time, such as the fibrinogen-erythrocyte binding [1]. For detailed methodology and some previously obtained results please see [1]. Briefly, this methodology involves: i) the covalent attachment of fibrinogen to the tip of the AFM; ii) the adsorption of blood cells on a poly-L-lysine treated glass slide (in buffer); iii) the taping with the fibrinogen treated tip until the occurrence of a binding event; and, iv) the pulling of the cantilever/tip, measuring the force necessary to break the fibrinogen-receptor bond.

2.2 Genetic studies

The GT patients genomic DNA is extracted from peripheral blood samples using standard procedures. The ITGA2B and the ITGB3 genes, which are associated with Glanzmann thrombasthenia due to the expression of the αIIb and β3 subunits, respectively, are screened for mutations. Intronic primers were designed to flank each of the 30 ITGA2B and 15 ITGB3 exons (GenBank accession number NM_000419 for ITGA2B and NM_000212 for ITGB3).

3 Results

Our previous results showed a significant reduction of fibrinogen binding both to platelets and erythrocytes on a GT patient with a β3 gene mutation, which indicates that the receptor for fibrinogen in erythrocytes has one of its units expressed by the same gene that is mutated in that patient [1]. To complement these results we studied patients with specific mutation(s) in the αIIb subunit of the αIIbβ3 receptor and genetically, biophysically and clinically characterized it. With these mutations we could obtain differences on the percentage of effective binding/unbinding events and on the force-rupture value between erythrocyte and fibrinogen, which can be very useful to conclude about the type of the specific fibrinogen erythrocyte membrane receptor. From the results we could observe that:

- If a GT patient has a mutation on the β3 gene (homo- or heterozygous), he/she has a significant reduction both on the frequency of fibrinogen-platelet binding and also on its force.
- If the GT patient has a mutation on the αIIb gene (homozygous only), he/she has a significant reduction of the frequency of the interactions between fibrinogen molecules and platelets, but not on its strength.
- If the patient has a mutation on the αIIb gene (also independently of homo- or heterozygosity), the force of the erythrocyte-fibrinogen binding is not significantly affected.
- The presence of a homo- or heterozygous mutation on the β3 gene leads to a significant decrease of the percentage of (un)binding events for the interaction between fibrinogen and erythrocyte and on its strength.
- The results reinforce the conclusion that the receptor for fibrinogen on the erythrocyte membrane is a product of the expression of the ITGB3 gene (β3 subunit). The findings obtained with these patients and their relatives suggest that the ITGA2B gene (αIIb subunit) has no interference on the fibrinogen-erythrocyte binding.

4 Conclusions and future perspectives

AFM force spectroscopy results correlated with the clinical and genetic data demonstrate the applicability of AFM as a highly sensitive nanotool for the functional evaluation of the outcome of genetic mutations resulting in hematological diseases. The differences on the fibrinogen-erythrocyte AFM force spectroscopy data (percentage of binding events and force) of patients with β3 or αIIb gene mutations, allow us to conclude about the characteristics of the erythrocyte membrane receptor for fibrinogen. To further prove that AFM force spectroscopy can be an excellent method of diagnostic of some diseases, we need to enlarge our studies and involve as many GT patients as possible.

Acknowledgements

The authors thank Mrs. Teresa Freitas (Faculdade de Medicina da Universidade de Lisboa) for excellent technical assistance. This work was supported by Fundação para a Ciência e Tecnologia – Ministério da Ciência, Tecnologia e Ensino Superior (FCT-MCTES, Portugal) project PTDC/QUI-BIQ/119509/2010.

References

Abstract

In 802.11 Wireless Networks, detecting faulty equipments, poor radio conditions, and changes in user behavior through anomaly detection is of great importance in network management. The traffic load and users’ movements on different access points in a wireless covered area vary from time to time, making these network management tasks harder. One of the objectives of this work is to inspect the evolving structure of wireless networks and their inherent dynamics in order to provide robust models for anomaly detection. For this purpose we propose to explore simultaneously the spatial proximity of access points as well as temporal usage patterns. The early phase of this approach consists of constructing a comprehensive model from the usage data of WLAN Access Points. We prepared a mixture estimation model of HMMs to achieve an optimized set of groups of APs. To evaluate our model we performed two other well-known clustering techniques, k-medoids and hierarchical clustering method and compared the total likelihood of the models with each other. The mixture of HMMs outperformed the other two clustering methods.

1 Introduction

As the employment of wireless infrastructures becomes an important part of peoples’ lives, utilizing communication devices has the potential ability of reflecting the lifestyle of human beings or the specific features of their inhabitant places. Understanding the user’s behavioral patterns will also play an important role in dealing with network management issues to provide higher quality of services, preventing network failures due to congestion, designing behavior-aware applications, and so on.

The main idea of applying Probabilistic Graphical Models and specifically Hidden Markov Models for such dynamic networks comes from the intrinsic similarity of the wireless hotspots fields with images, as the pixels or blocks of images remind the APs or APs’ bundles in a WLAN. Moreover, the temporal sequence of data given for every spot resembles the succession of images in a video or in higher dimensional images. Hence, as the early steps towards the higher dimensional HMMs, the mixture of HMMs [3], [9] have been employed to attain a robust set of clusters.

The mixture estimation technique behaves very similar to the mixture of Gaussian method for the clustering purpose, with distinction of performing simultaneously the spatial proximity of access points as well as temporal usage patterns. The mixture probability density function (pdf) is parameterized by a set of parameters that describe the density functions of linear HMMs with multivariate emission distributions. The observed data $O_t$ is modeled to a mixture of $K$ hidden states $\lambda_k$, $k = 1, ..., K$, and the likelihood function of the HMMs denoted by $P(O_t|\lambda_k)$. The $\lambda_k$ is the set of parameters that describe the density functions of linear HMMs with multivariate emission distributions. The observed data $O_t$ then corresponds to the multi-dimensional time-courses that reflect the usage pattern of the APs. We now aim at maximizing Equation 1 by choosing optimal parameter set. This problem is generally solved by the EM algorithm which finds a local optimum for the above function.

The mixture method concisely performs the following main steps, given a collection of $K$ initial HMMs $\lambda_1^1, ..., \lambda_K^1$:

1. **Iteration:**
   - Generate the initial groups of sequences by assigning each sequence $O_t$ to the model $k$ for which the likelihood is maximal.
   - Calculate new parameters for each model $\lambda_t^1, ..., \lambda_t^K$ using re-estimation algorithm based on their current parameters $\lambda_t^{k-1}$ and the participating sequences and their assigned weights.

2. **Stop:** if the improvement of the objective function is below a given threshold $\epsilon$, the grouping of the sequences does not change or a given iteration number is reached.

3 Data Set

The raw data set of this paper consists of the daily summary of connections between 211 access points and their corresponding stations in one of the busiest Eduroam hotspots at the Faculty of Engineering of the University of Porto (FEUP). The data used for the current work belongs to a 5 month period of tracing from 2008-11-11 to 2009-04-11. The entire data set incorporates records of 802.11 mobile stations’ association to APs stored at a RADIUS authentication server. When a client associates/disassociates to an 802.11 AP, a Start/Stop event is recorded. A probing log event Alive is provoked every 15 minutes while the client is still connected to the network [1].

3.1 Features

A number of features emerge from the raw dataset as a result of an enumeration process with application to model building and data analysis.

**User Count:** the number of unique users observed in a specific location (indicated by an AP) in a temporal period.

**Sessions:** the raw count of active sessions during a time-slot regardless of the owner. This attribute shows the number of attempts has been made to associate with an AP by a number of users.

**Sightings:** the probing event occurring every 5 minutes as an indicator of users’ duration of stay in a location, and a counter for sessions longer than 5 minutes.
Table 1: Total likelihood Results of Mixture of HMMs, K-medoids and Hierarchical Clustering Techniques.

<table>
<thead>
<tr>
<th>Clustering Methods</th>
<th>Total Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixture of HMMs</td>
<td>-37318.59</td>
</tr>
<tr>
<td>K-medoids</td>
<td>-551926.50</td>
</tr>
<tr>
<td>Hierarchical Clustering</td>
<td>-571260.53</td>
</tr>
</tbody>
</table>

4 Experimental Results

From the data sequences of N APs, we build one model for each regarding the duration of 5 working days and 11 idle hours per day. The reason we bypassed the weekend and the night hours is that there is not adequate movements during these time spans, and the sparse data set is not very advantageous for the further calculations. Thus given the three features of user count, sessions and sightings, for the mentioned period of 55 hours, the HMM models are built with multivariate Gaussian emission distribution. Each HMM has 3 states, indicating the low, medium and high status of AP’s temporal usage, and these 3 states are the result of a vector quantization process on the data set (3 clusters produced by k-means). This preliminary clustering step, improve the HMMs’ initial parameterization and provide more reliable models to begin the EM iterations with.

Once the models are constructed for every AP; the mixture estimation process is applicable on the sequences. However, not all the models are dominant enough so that their pdf could provide the likelihood result of all or majority of sequences. So a number of proficient models are selected by a filtering process and the final mixture model is constructed on a subset of these models. Due to the random factor of the algorithm, the entire procedure is repeated 20 times and the model with the maximum likelihood is selected.

4.1 Evaluation

To evaluate the excellence of the formed clusters by mixture of HMMs, we apply a number of well-known clustering algorithms to the same set of data making different groups of APs. The total likelihood of each model is calculated aggregating the likelihood of each cluster within the model. And the likelihood of each cluster is obtained by making the minimum spanning tree of the participating members of the cluster. Many other methods are also applicable to attain a likelihood value for each cluster, for instance accumulating the weights of all the edges in a complete graph of each cluster etc. However as the likelihood in this context implies dissimilarity of the data sequences, we selected minimum spanning tree to estimate the best case likelihood and thus the overall minimum distance of each cluster.

To perform other clustering techniques, we need to provide a distance matrix to the clustering algorithm. Similar to the mixture estimation, k-medoids and hierarchical clustering methods are also re-iterated, each 500 times, to reach to their most optimized solution with the minimum error. When the groups have been created using one of these clustering techniques, the sum of likelihoods of each group give the total likelihood of the model, hence the grouping methods can be compared.

There are several approach to construct the distance matrix, presenting a notion of similarities/dissimilarities between the data sequences, and specifically in this work, AP’s behavioral patterns. The symmetric distance of two sequences $S_i$ and $S_j$ is defined as:

$$d_{ij} = \frac{1}{2} \left( P(O_i|\lambda_j) + P(O_j|\lambda_i) \right)$$

As indicated in [5] the above equation do not take into consideration the quality of the models themselves, how well sequence $S_i$ is modeled by its own HMM $\lambda_i$. Regarding this issue, another distance equality is defined which contemplate the capability of models in generating their own sequence as well:

$$d_{ij} = \frac{1}{2} \left( \frac{P(O_i|\lambda_j) - P(O_i|\lambda_i)}{P(O_i|\lambda_i)} + \frac{P(O_j|\lambda_i) - P(O_j|\lambda_j)}{P(O_j|\lambda_j)} \right)$$

Having constructed the distance matrix, the total likelihood of the entire data sequences for every clustering technique is demonstrated in Table 1. As expected the number of clusters in all the mentioned techniques is equally set to be the same as the number of clusters (or mixtures) in the mixture estimation method. Total likelihood of all the models including mixture of HMMs is negative. As the length of the observation emissions are relatively high, probability of the observation sequences are very small numbers and then the log operation on them return negative values. In spite of the modification made by Equation 3 on the sequences’ distances, and although positive total likelihoods are also probable, the negative result is not unanticipated at all. However, the outcome of the mixture of HMMs approach is less negative than the two other methods and it shows a higher total likelihood compared to k-medoids and hierarchical clustering results.

5 Conclusion

In this paper we mainly focused on the cluster formation of the wireless APs using mixture of HMMs. This work is among the early steps of performance anomaly detection in 802.11 wireless networks for the network management purposes. A 3-state HMM is built for every participating AP and then the preeminent models are selected to take part in the EM algorithm of the mixture estimation. The final converged mixture model produce a number of mixtures (or clusters) and one optimized HMM per cluster. The total likelihood of the entire data set is measured by aggregating the clusters’ likelihoods which derive from the sum of edges of the formed minimum spanning trees. The proposed clustering approach is evaluated with 2 other well-known clustering techniques, k-medoids and hierarchical clustering. The total likelihood of the clusters generated by the mixture model outperforms the two other techniques.

A relevant direction of future work is to perform constraint mixture estimation [2] in order to involve the physical proximity of the APs [8] to the clustering result as well. Next step contains the anomaly detection, observing the amount of divergence of the hourly likelihood outcomes and make a comparison between one single model for all the APs, one model per AP and one optimized model per group of APs (mixture model).

Acknowledgment

This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project Place Characterization from Sensing and Acting: EXPL/EEI-SII/0902/2012 (FCOMP-01-0124-FEDER-029831).

References

**Abstract**

Some human pathologies such as Candidiasis have become resistant to common antifungal drugs. Psd1 is an antifungal peptide with antifungal activity against this pathogen. Based on atomic force microscopy technique some parameters such topology and stiffness were evaluated. Obtained results show a correlation of these parameters with cell death, membrane disruption and leakage of cellular content.

1 Introduction

Psd1 is a defensin, isolated from Pisum sativum seeds, previously shown to have a strong interaction with fungal-specific membrane components [1]. Candida albicans is an important human pathogen, causing oral, genital and systemic opportunistic infections, which are especially relevant clinically in immunocompromised patients, such as HIV-infected individuals. We tested the effects of this antimicrobial peptide and two well known antifungal drugs amphotericine B and fluconazole, at the minimal inhibitory concentration (MIC) and at a 10-fold higher concentration. By atomic force microscopy (AFM) imaging we assessed morphological changes on C. albicans cells. Our results show that, with increasing incubation times and Psd1 concentrations, there is an increased cell death and surface roughness, with the appearance of apoptotic features, such as membrane blebs, cell size alterations, membrane disruption and leakage of cellular contents. Thus, we were able to visualize the action of Psd1 against a relevant fungal human pathogen, aiming at its possible use as a natural antimycotic agent.

**References**

Towards efficient path planning of a mobile robot in rough terrain

Diogo Amorim
diogo.amorim@ist.utl.pt

Rodrigo Ventura
rodrigo.ventura@isr.ist.utl.pt

Abstract

Path planning for autonomous vehicles on rough terrain is a different problem to solve comparing to traditional 2D, flat terrain path planning. We address this problem by first constructing a cost grid map where each cell represents a passage/progress cost for the robot rather than, for example, an occupancy grid map given a known surface representation. The cost value for each \((x,y)\) is based on the pose of the vehicle when on the surface and it results from the pitch angle. One possible way of calculating the robot’s pose is to define that as a constrained optimization problem with non-linear constraints.

1 Introduction

This paper shows the steps being taken in order to create a path planning method for a robot presented with a rough terrain. Path planning for 2D maps is usually based on an occupancy grid map with a \((x,y)\) configuration where each entry depicts whether or not that position is free to the robot [7], and the total path cost is usually the distance until the goal position or some simple variation of that value. There already are some path planning methods that solve 2D map path planning problems which are powerful tools [6], and so, part of the intention of this project is to apply these tools to the same sort of problem, and with the same goal, but with a different premise. This premise is the type of map, a map that represents a rough surface has information about the free space and insuperable obstacles but also the elevation of each coordinate. The elevation variations can implicate new obstacles, i.e., if a slope if too steep the vehicle will not be able to climb it.

The purpose of this paper is to show how to create a map in a \((x,y)\) configuration where each coordinate is associated with a cost based on the vehicle’s pose if it were to rest on those coordinates. Defining the problem as a constrained optimization problem with non-linear constraints it is possible to determine the pose of the mobile robot as well as the number of contact points with the ground, an important factor to determine whether or not it is possible to stand on that position.

After creating a map of the robot’s world depicting these degrees of difficulty one could, in principle, apply a number of path planning methods already in existence. The final goal of this project is to develop a controller for the RAPOSA-NG 2, a track wheel robot designed for urban search & rescue operations, that can be equipped with depth sensors. The controller will efficiently drive the robot from point A to point B in any rough terrain.

2 Problem Statement

At this stage we are still at a proof of concept stage where the environment is all simulated. The map is in a \((x,y,z)\) configuration, defined in the inertial frame of reference \((i)\) and is derived from a sinuosoidal surface easily obtainable from the mpl_toolkits package for python. The resulting surface has a bump like shape with a smooth slope and a sharp end as shown in Figure 1. The robot is defined as a 7 point \(p_{ij} = (x_{ij},y_{ij},z_{ij})\) structure Figure 1, all fixed to its reference frame, the robot’s reference frame \((r)\) where the origin is set at the center of mass of the robot. Three points characterize each contact point of a track, one represents its beginning \((p_{1r},p_{2r})\), another its end \((p_{3r},p_{4r})\) and another its middle point \((p_{5r},p_{6r})\). The last point \((p_{7r})\) represents the vehicle’s center of mass, which, as previously mentioned, was described as the origin of the frame.

The robot’s pose is defined as:

\[
(x,y,z)_{ij} = (x_{ij},y_{ij},z_{ij})
\]

where \((x_{ij},y_{ij},z_{ij})\) are the coordinates of the \(r\) in relation to the \(i\) and \(\theta\) is the angle of rotation of the \(r\) around the \(Z_r\) axis, \(\beta\) the \(Y_r\) axis and \(\gamma\) the \(X_r\) axis. To determine the coordinates of the points defining the robot on the \(i\) their coordinates in \(r\) are multiplied by a rotation matrix:

\[
\begin{bmatrix}
\cos(\beta) \cos(\theta) & \cos(\theta) \sin(\gamma) \sin(\beta) & \cos(\gamma) \sin(\beta) - \cos(\beta) \sin(\theta) \\
\cos(\beta) \sin(\theta) & \cos(\theta) \cos(\gamma) \sin(\beta) & \cos(\gamma) \cos(\beta) + \sin(\theta) \sin(\beta) \\
-\sin(\beta) & \sin(\theta) \cos(\gamma) & \cos(\gamma) \cos(\beta) + \sin(\theta) \sin(\beta)
\end{bmatrix}
\]

and then adding the position of the \(r\) relative to the \(i\).

The target is to minimize the \(z\) coordinate of the robot’s center of mass, in relation to the \(i\) but with the restriction that none of the points that define the robot can pass through the surface. The vehicle also poses limitations, it cannot be upside down and cannot climb hills steeper than 45 degrees (value stipulated for the simulation) so its roll and pitch angle absolute maximums were set at that same value. These limitations are translated as contrications when inserted in an optimization problem, and so, in order to solve this specific problem one can resort to a constrained (multivariate) problem solving routine already available and developed. In order to introduce constraints in the optimization function it is necessary to formulate them as inequalities, functions whose values are always positive which in this case means, for example, that the \(z\) coordinate of each point defining the robot minus the \(z\) coordinate of the point of the map directly below must be positive, and this condition is respected by the algorithm. There are 9 constraints to this simple problem, one per each point that defines the robot, one for the roll and another for pitch angle. More contrications can and will be added in order to simulate the hull of the vehicle more accurately simulating it. The task of determining the robot’s pose can then be defined as a constrained optimization problem in the following way:

Minimize:

\[
z_{ij}
\]

Variables:

\[
z_{ij}, \beta, \gamma
\]

Subject to:

\[
z_{ij} - map_{x_{ij},y_{ij}} \geq 0
\]

\[
\frac{\pi}{4} - |\beta| \geq 0
\]

\[
\frac{\pi}{4} - |\gamma| \geq 0
\]

The optimization function minimizes the value of \(z_{ij}\) by manipulation of the three variables it has access to: \((z_{ij},\beta,\gamma)\). The contrications determine that the robot’s pitch (\(\gamma\)) and roll (\(\beta\)) angles don’t reach values greater than 45° (\(\frac{\pi}{4}\) rad) or smaller than \(-45°\) (\(-\frac{\pi}{4}\) rad). Another limitation is that none of the \(z\) coordinates of the points defining the robot \((z_{ij})\) can be lower than the elevation of the map directly bellow \((map_{x_{ij},y_{ij}})\) thus \(z_{ij} - map_{x_{ij},y_{ij}}\) must be greater than 0. The results are as expected, the robot touches the ground with three or more of the six points defined as its tracks depending on the surface roughness. All those are valid positions, but if the function returns that only two or less points are touching the surface that means that the autonomous system can’t be on that position because of pose limitations introduced as constraints and that same position on the map is considered an obstacle. It is now possible to build a new map where instead of \(x\) coordinates we use a combination of the pose angles (pitch and roll) provided by the previous routine. For every cell of the map, i.e., every time the optimization function returns a pose, the value of \(\gamma\) is stored in the equivalent cell of the cost map. If we were to keep the absolute value of \(\gamma\), we would be admitting the cost of moving uphill or downhill with equal inclination is the same. It is intuitive to say the robot will struggle more going uphill then going downhill and this is why we kept the information about the sign of \(\gamma\). We now have the information about where the map, at a certain \(\theta\), is up or downhill. At the end of the process of calculating the pose of the robot for every map cell, the lowest cost value is found and added to every cost map cell, this action prevents the existence of negative cost values. For each possible \(\theta\) angle of the robot a different cost map will be produced by the algorithm, and so there is not just one 2D matrix depicting the cost map at all.
as the possible θ angles the robot can assume. All these matrices can form a 3D matrix where each layer is the map for a specific θ angle.

3 Results

The process explained above was applied to the already mentioned bump like surface. This map of 1225 cells takes about 0.013 seconds/cell to be processed but there is still room for speed improvement as we are still developing concepts.

As we can see from the contour map on Figure 2 (a) of the bump the cross sections are almost tear shaped and this explains the behavior of Figure 2(b). The cost map obtained for θ = 0 from the algorithm is shown in Figure 2(b), it shows cost values ranging from dark blue to dark red as shown on the side bar. Higher cost values coincide with steeper uphill slopes and lower cost figures translate steeper downhill slopes. This cost map is only applicable to a robot travelling bearing θ = 0, so one could imagine a robot starting from (0, −30) (bottom of the graph) and travelling to (0, 20) or in lines parallel to that. The light green/yellow line demonstrates the beginning of the slope, until then the cost was constant, but now the robot is climbing uphill the cost values will increase. The values go from green/yellow up to red (if the color does not change it means the map has a constant inclination). Because the surface is rounded in all edges the transition from uphill to downhill is translated by the green/light blue line, and as the surfaces becomes steeper (downhill) again the blue color darkens. At the end of the descent the color goes from light blue to green and the cost represented by the green color is the same up to the upper edge of the map indicating, again, a constant slope. The Scipy package for python includes a function (fmin_cobyla) that solves Constrained Optimization problems BY Linear Approximation (COBYLA) [4] and it is one of the possibilities that was used to obtain the results here shown. The fmin_cobyla respects the constricitions to a user defined error margin which here was defined as a millimetre (although the authors do not guarantee 100% accuracy in every constraint, the results for the tested surfaces were always within the defined error).

4 Conclusions and Future work

Although being a different path planning problem, the presented conundrum can be solved with adaptations of technologies already in existence. Simulating the structure of the robot and using an elevation map it is possible to create a cost map, by computing the robot’s pose in every map cell. This cost map is a crucial tool for the task ahead, path planning. We are still in an early stage of the solution and because of that, some questions that arise when solving the main problem are not completely answered yet. The future work will be about the use of the cost map as an input for an already existing path planning method. It is theoretically possible to apply a Fast Marching Method (FMM) [2, 5], rapidly exploring random trees (RRT) [1, 3] or other path planning methods and obtain an efficient path for the robot.

References

Abstract

In this work we propose a methodology to find automatically the type of the lens of a discrete mobile camera. The assumption that pixels have approximately uniform density on the sensor allows the classification of different types of lenses, independently of the sensor shape.

1 Introduction

Traditional imaging sensors are formed by pixels precisely placed in a rectangular grid, and thus look like calibrated sensors for many practical purposes such as localizing local extrema, edges or corners. In contrast, the most common imaging sensors found in nature are the compound eyes, collections of individual photo cells which clearly do not form rectangular grids, but are very effective for solving various tasks at hand and thus have inspired the design of many artificial systems.

Recently, Olsson et al. [8] proposed a methodology for topologically calibrating a central imaging sensor based on a number of photo-cells. A metric reconstruction is found by Grossmann et al. [5], when the relation between signal correlation values and pixel distance-angles is known. Methods that do not require this relation to be known were presented by Censi and Scaramuzza [2] and Galego et al. [6]. In [6] the computational complexity associated to augmenting the sensor resolution is handled by using methods derived from the classical Multi Dimensional Scaling (MDS) [3].

In the cases where the sensor topology is a rectangular grid with a perspective lens one can use traditional calibration methodologies [1, 9, 11]. However, for other type of lenses these methodologies do not work. A methodology to calibrate other types of lens was proposed by Kannala et al. [7]. The methodology of Kannala et al. requires using a calibration pattern and the specification of the lens type. In our work we propose an automatic method to find the lens type while using natural images.

2 Camera Model

Discrete central cameras, as conventional (standard) cameras, are described geometrically by the pin-hole projection model. Differently from standard cameras, discrete cameras are simply composed of collections of pixels organized as pencils of lines with unknown topologies.

Grossberg and Nayar [4] introduced the concept of raxel as a mathematical abstraction of the pose of a photo-cell. Instead of denoting the real position of the photo-cell, a raxel is just assumed to be along the direction of the chief ray associated to the photo-cell. A raxel can be characterized as a 3D position, p, and a direction vector, q, as shown in Fig.1(d). Since we are considering central cameras, all light rays (acquired by photo-cells) converge to the same point, and thus p1 = p2 for any pair of raxels. Therefore, we ignore the position of the raxels, since the only useful information is contained in the direction vector q. As a direction vector, q, we assume that all the vectors have the same norm. This assumption removes one degree of freedom, which allows us to represent q with only two angles, (Ω, μ).

Traditionally the coordinate system of a camera sensor is represented by u1 and v1, which are the ith pixel position along horizontal and vertical grid. Here we use a polar coordinate system of [r μ], where r = \sqrt{(u1 - u0)^2 + (v1 - v0)^2}, μ = arcos(u1/v1), and [u0 v0] is the principal point.

In this work we assume that the discrete camera geometric model can be characterized by an unknown radial function h. This function links the angle at which a light ray (raxel) hits the camera lens with the imaged point (pixel coordinates),

Ω = h(r/|l|)

(1)

considering that r is the radial distance, in pixels, from the center of an imaging sensor, l is the focal length and Ω is the angle between the principal axis and the incoming ray, as it can be seen in figure 1 (d), note that the μ is the same as in pixels coordinates, since the lens transformation only affect the radius.

In the following we assume that we have three different lenses [7]:

\[ \Omega = \text{atan}(r/l) \] perspective lens, \hspace{1cm} (2)
\[ \Omega = r/l \] equidistance projection lens, \hspace{1cm} (3)
\[ \Omega = \text{asin}(r/l) \] orthogonal projection lens. \hspace{1cm} (4)

From now on the focal length l will be not considered since it is a constant that will not have impact on the differentiation of a lens type.

In order to classify a lens mounted on a camera we propose a methodology based in three steps: i) topological calibration of a sensor; ii) marginalization of the density of the topology along μ; iii) lens classification based in finding the closest match for the marginal density function of the topology.

3 Auto-Calibration Methodology

The classical Multiple Dimensional Scaling (MDS) algorithm [3] provides a simple way of embedding a set of points in Euclidean space given their inter-distances. It works well when the distances are Euclidean and when the structures are linear; however, when the manifolds are nonlinear, the classical MDS fails to detect the true dimensionality of the data set. Isomap is built on classical MDS but instead of using Euclidean distances it uses an approximation of geodesic distances [10]. These geodesic distance approximations are defined as a series of hops between neighboring points in the Euclidean space using a shortest path graph algorithm such as Dijkstra’s. In our particular case, this algorithm is used to provide a pixel embedding given the inter-pixel distances estimated from the pixel stream correlations.

In order to obtain the embedded raxels directions, Q = [q1 q2 ... qn], one follows the steps proposed in [6]: (i) Data binarization using a fixed threshold such that each pixel stream value is either 1 or −1. (ii) Computing the normalized correlation between all the pixel-streams. (iii) Converting the inter-pixel correlations, C, into distances, d, using the linear transformation \( d(q_i, q_j) = 1 - C(f_i, f_j) \), \( f_j \) corresponds to a time series of brightness values captured by \( i^{th} \) pixel. (iv) Using Isomap to compute the topology of the sensor.
The truncation is done since a random topology is in general not circular, cumulative functions, are made using the biggest circle that could fit in the topology. If the first term has a modulus value lower than 10% of the lens mounted in our camera. The first case that we look at is the one of the camera, 100×100 pixel camera with the three types of lens one finds that each lens creates a distinct distribution of raxels (see Fig. 3(d)) 1.

Given the previous observation one can now propose a lens classification algorithm. One starts by acquiring the radial distribution of the topology, \( h^{-1}(\Omega) \). We truncate the domain of \( h^{-1} \) to \( \Omega \in [0, h(\max(h^{-1}(\cdot)))] \). The truncation is done since a random topology is in general not circular (e.g. most of the conventional sensors have rectangular shapes). Then a quadratic curve is fitted to \( h \) (e.g. most of the conventional sensors have rectangular shapes). Then a quadratic curve is fitted to \( h \). The part of the function used for classifying the lens is marked in red. The estimated quadratic curve is \( y = 0.0045 x^2 + 0.19 x + 0.013 \). In this case the ratio between the first and the second term is 2.3%, which means that the lens mounted in the camera is an equidistant lens. This result is confirmed by the datasheet of the lens.

Figure 2(d) shows the density function of the estimated topology. The marginal density of the raxels estimated quadratic curve is \( h \). Part of the function used for classifying the lens is marked in red. The estimated quadratic curve is \( y = 0.0045 x^2 + 0.19 x + 0.013 \). In this case the ratio between the first and the second term is 2.3%, which means that the lens mounted in the camera is an equidistant lens. This result is confirmed by the datasheet of the lens.

Figure 3: Effects of a lens in a uniform pixel distribution to the raxels topology. (a) Perspective lens. (b) Equidistant lens. (c) Orthogonal lens. (d) Density of each topology along the radius.

4 Lens Effect and Raxel Densities

Assuming that we have a uniform distribution of the pixels, what is the distribution expected in the raxels space? The marginal density of the pixels would be the solution if we had no lens transformation, however the transformation created by the lens will change the marginal density of the raxels.

Simulating a 100×100 pixel camera with the three types of lens one finds that each lens creates a distinct distribution of raxels (see Fig. 3(d)) 1.

According to the reconstructed topology. Since the estimated topology is not perfect, and we have pixel sub-sampling one can see blur in the figure. Figure 2(d) shows the density function of the estimated topology. The part of the function used for classifying the lens is marked in red. The estimated quadratic curve is \( y = 0.0045 x^2 + 0.19 x + 0.013 \). In this case the ratio between the first and the second term is 2.3%, which means that the lens mounted in the camera is an equidistant lens. This result is confirmed by the datasheet of the lens.

6 Conclusions and Future work

In this work we have shown that is possible to find automatically the type of a lens mounted on a mobile camera. This is useful, for example, to further automate current calibration processes which involve indicating a-priori the type of the lens. Our future work will focus on formalizing the mathematical background of the proposed methodology.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE/EEI/LA0009/2013, by the FCT project PTDC/EEA/105413/2008 DCCAL, and by the FCT project EXPL/EEI-AUT/1560/2013 ACDC.

References

Abstract

In order to improve the players’ and teams’ performances in sports, the technical staffs are now using an ever increasing number of technologies. In this paper we present a study based on a video tracking of players and ball, that can return heat, passes and ball losses maps, which can be a good resource to measure a player/team performance and a base to their improvement.

1 Introduction

Paper support maintains until our days as one of the basic ways to log the player’s and team’s actions (e.g., passes, rebounds, turnovers, recurrent movements). At the end of the matches, those documents are statistically analysed by the technical staff and the players are notified about their actions. The log work is made by someone that is observing the game (live or recorded) and marking those moments by hand. This is a slow process and, in live games, can require more than one person to process all actions with an optimum precision.

This paper shows our alternative to that kind of work. Our system rests on a one camera tracking system: (a) a portable camera that can be placed in the stands to acquire the images; (b) software to process the (live or recorded) matches. The returned data is then used to show individual or team actions. This work deals with the constructions of the heat, passes and loss ball maps for a football match. With this process, the information can be given in real time, allowing an instantaneous analysis of the players and teams actions, which gives the possibility to the technical staff to make in game adjustments, based on that information. As expected, knowing the opposite team performance is also possible with this system, which can give an extra advantage to those using these features. This work is a part of the Footdata project [1].

On the market there are some systems which incorporate partially some of the components of this project, for example, Kizanaro [2] Tracab @ [3] Prozone [4] Amisco [5] and SportVU [6]. We can also find some research that explores aspects of the project, such as Spatial and Spatiotemporal Analysis of Soccer [7], and tactical analysis in football, [8], [9] and [10]. Our system, which integrates the tools presented in this paper, will integrate some of the functionalities that these systems already have.

Since one of the main objectives of the Footdata project was to build an operating system (OS) independent software it was decided to make it as a web application, so it can run in all equipment that have a browser (such as, Chrome, Firefox, or Safari) and an internet connection. The web application is being developed with the django framework were all the coding is made in python

HM is build using the data from the tracking system, that returns for each frame the players’ and ball’s positions in meters, 

\[
\text{HM} = \begin{pmatrix}
\text{frame_id} & t & \text{teamA}' & \text{playerA} & \text{x1_A} & \text{y1_A} & \text{dx_A} & \text{dy_A} & \text{v_A} \\
\text{frame_id} & t & \text{teamB}' & \text{playerB} & \text{x2_B} & \text{y2_B} & \text{dx_B} & \text{dy_B} & \text{v_B}
\end{pmatrix}
\]

Those positions are relative to the pitch top left corner and are obtained using an homography [14]. To compute HM we start by setting it to the null matrix, \(HM = [0]\), then for each frame a cross-multiplication is used to get the corresponding positions of the players in the pitch, \((x, y)\), to their corresponding entry in HM matrix, \((x', y')\), and HM_{opt} is incremented by one unit. Once all the frames and players are processed, the matrix is normalized, i.e., each entry of the matrix is divided by the matrix maximum element. Finally, the HM values (ranging from 0 to 1) are used to build the SVG image, by setting the colour opacity of the pixels that form the heat map.

Figure 1 sketches a team heat map from one minute of a football match. The heat map can be filtered within different time intervals, and individual or multiple players (Figure 2).

3 Passes Maps

The passes maps are a graphical representation of the passes made by the players in the game. It shows the passing player’s number and the receiver’s number. The passes map is a tool which allows several analyses, like how a team organises the attack or maintains the ball possession. Other statistics can also be inferred like who are the most important players in the teams’ strategies, in the sense that they have more ball actions.

The following process is operated to find the passes between the players. In each frame, (1) the distance between the players and the ball is calculated; (2) if the ball enters in a player’s radius (empirical tests, revealed good results considering a radius distance equivalent to 1.3 meters), that frame and player are tagged, and the ball velocity and direction are calculated using the current position and its position five frames before; (3) if the ball leaves that player’s radius, the velocity and direction are also calculated, considering the ball’s current position and its position five frames after; (4) the velocity and
direction values are then subjected to a set of conditions (changes in the ball direction or velocity), in order to see if it’s a pass or not: if the ball changes direction or velocity then it is almost sure a pass.

The pass detection process is summarized in Algorithm 1. Figures 3 - 4 show a passes map where it’s possible to see all the passes made by both teams, a team or by a single player.

Simultaneously, a log is built as the game is processed, containing a list with the intervening players numbers, teams and time of the pass execution and reception.

Algorithm 1: Passes Extraction

For each frame

If ball enters player radius

\[
\begin{align*}
vi &= \text{velocity(frame\_id, frame\_id - 5)} & // \text{ball initial velocity} \\
di &= \text{direction(frame\_id, frame\_id - 5)} & // \text{ball initial direction}
\end{align*}
\]

If ball leaves player radius

\[
\begin{align*}
vf &= \text{velocity(frame\_id, frame\_id + 5)} & // \text{ball final velocity} \\
df &= \text{direction(frame\_id, frame\_id + 5)} & // \text{ball final direction}
\end{align*}
\]

If \(di <> df\) or \(vi <> vf\)

Almost sure it’s a pass

Else

Not considered as a pass

5 Conclusion and future work

In this paper we have shown a small part of the overall system that is being built to help coaches and technical staff in the improvement of their players and team’s performance, by using: heat maps, passes maps and loss ball maps. The maps can be filtered by time intervals, by team or by players, in order to in an easier way understand and evaluate the player’s / team performance. It was also explained our approach in extracting the passes from the information given by the tracking system and how we get the ball losses.

The system is being optimized for football, but the difference between other ball sports is not big, so maybe in a near future our system will be adapted to other sports.

Acknowledgements: This work was supported by FCT project PEst-OE/EEI/LA0009/2013 and project FootData QREN I&D&T, n.º 23119. We also thanks to project leader Inesting, S.A. [www.inesting.com], and the consultant football coach Domingos Paciência and our colleague Carlos Gomes.

6 References


Development of amyloid-based biomaterials for nanotechnology

Gabriela M. Guerra, Sónia Gonçalves, Nuno C. Santos, Ivo C. Martins

Instituto de Medicina Molecular (IMM), Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal

Abstract
Amyloid fibrils, often associated with human degenerative diseases (such as Alzheimer’s and Parkinson’s diseases), may also have physiological roles, having even been suggested as potential novel biomaterials. In general, due to their β-sheet rich architecture, amyloid fibrils have an exceptional stability, mechanical strength and resistance to degradation, rendering them excellent nanomaterial candidates. The potential to form amyloid fibrils (and other protein/peptide aggregates) can be predicted from the peptide amino acids sequence. Using atomic force microscopy (AFM), circular dichroism (CD) and Fourier transformed infra-red (FTIR) spectroscopy, we tested the ability of three amyloid peptide sequences (STVIIE, ISFLIF and GNNQQNY) to form amyloid fibrils and/or β-sheet rich protein aggregates at different times of incubation. AFM, CD and FTIR data, taken together, indicate the peptide STVIIE as the most reproducible and amenable peptide for developing amyloid-based nanotechnology and nanomedicine approaches.

1 Introduction
Amyloid fibrils are well-ordered structures often associated with several human degenerative diseases, such as Alzheimer’s and Parkinson’s diseases [1–4]. Due to their role in pathological processes, amyloid fibrils have been thoroughly studied in terms of amyloid kinetics, morphology, secondary structure and toxicity [1–5, 7, 8]. It is clear that amyloid kinetics, i.e., the time that amyloidogenic amino acid sequences take to form amyloid fibrils, is dependent on a number of key factors, of which the main ones are: primary sequence of the specific protein/peptide (with hydrophobicity, net charge and presence of specific gatekeeper residues that prevent/promote aggregate being propagated); 2) conformational change of the peptide (e.g., tendency to form unstructured aggregates) [3]; 3) protein/peptide concentration (high concentrations favour amyloidosis) [5]; 4) presence of seeding nucleus (which accelerate the process) [2, 3, 5]; temperature (higher temperatures promote aggregation and amyloidosis) [2–5] and, finally, specific buffer conditions [2–5]. Regarding morphology and secondary structure, it is clear that amyloid fibrils are highly ordered, quasi-crystalline β-sheet rich structures organized in a cross-β arrangement quite distinct from other forms of protein/peptide structures and heterogeneous aggregates [2, 3]. Toxicity of amyloid species is a highly interesting point since it is what triggered the original interest in these amino acid sequences and led to the careful characterization of the fibrils formation. Surprisingly, contrary to the original expectation, amyloid fibrils were shown not to be the major cause of toxicity in amyloid diseases [3–5]. In fact, mature amyloid fibrils are relatively inert and mostly innocuous, with toxicity being now attributed to the fibrils precursor intermediates, i.e., the oligomers and protofibrils formed early in the aggregation process [2–5]. The discovery that amyloid fibrils are not as toxic as once assumed was accompanied by the finding that there are a number of the so-called functional amyloids, which are peptides or proteins that, while in β-sheet rich amyloid fibril conformation, play a number of physiological roles, in several different organisms, from bacteria to invertebrates and even to mammals, including humans (reviewed in [1]). This elicited a new view on amyloidogenic species, especially from the fields of biomaterials and nanotechnology [1].

Nanotechnology advances require major developments in the field of biomaterials through the development of chemical species with rationally designed physicochemical and mechanical properties [1]. A major requirement is resistance to the surrounding environment, in order for chemical reactions to occur in their immediate vicinity without being affected by them [1]. Other major requirement is the ability of the material in question to be chemically modified for the function desired without affecting the above mentioned properties [1]. In many applications, self-assembling materials are also highly sought, especially those that organize themselves in a well-established manner in different and controlled topographies [1]. Amyloid fibrils possess these characteristics and are for this reason promising biomaterials [1].

Overall, the mature amyloid fibrils architecture, a series of β-sheet structures that are composed of hydrogen-bonded β-strands running perpendicularly to the fibril axis [1–3], is a highly promising biomaterial especially if these mature fibrils are stable at physiological buffer conditions of pH and temperature. Building on that knowledge and based on previous work [4, 5, 7, 8], we tested the ability of three peptide sequences (STVIIE, ISFLIF and GNNQQNY) to form amyloid fibrils and/or β-sheet rich protein aggregates at physiological buffer conditions of pH and temperature, aiming to obtain a variety of behaviours suitable for different applications in nanotechnology and nanomedicine.

2 Materials and Methods
To induce fibril formation, lyophilized peptides (JPT Peptide Technologies GmbH, Berlin, Germany) were first dissolved in 50 mM Tris-HCl, 5 mM EDTA, pH 7.5 at different final concentrations (STVIIE and GNNQQNY at 1 mg/mL and ISFLIF at 0.1 mg/mL) and bath sonicated for 5 min to disaggregate pre-formed peptide aggregates. Amyloid beta peptide 42 (Aβ42), a widely studied amyloid forming sequence [1–4], was employed as a positive control, being incubated at 1 mg/mL in the same conditions. The solutions were then incubated at room temperature and fibril formation was allowed to proceed. At different time points (0 h, 24 h and 2 weeks), the presence of amyloid fibrils and its secondary structure content were studied by atomic force microscopy (AFM), circular dichroism (CD) and Fourier transformed infra-red (FTIR) spectroscopy. CD measurements were recorded on a Jasco spectropolarimeter J815 using quartz cuvettes (Hellma) with path lengths of 1 mm. A scan rate of 200 nm/min was used and 5 spectra were averaged for each measurement. Samples were thermostated at 25 °C using a Peltier system connected to a recirculating water bath. FTIR measurements were performed using 40 µL of incubated sample in a Bruker Tensor 27 equipped with a BioA TR flow cell. AFM images of the dried samples immobilized in poly-L-lysine slides were acquired in air, using a JPK Nanowizard II in intermittent contact mode and App Nano ACL50 cantilevers with a spring constant of 58 N/m and a frequency of 190 kHz.

3 Results and Discussion
The ability of the three peptide sequences (STVIIE, ISFLIF and GNNQQNY) to form β-sheet-rich amyloid fibrils was confirmed using AFM, CD and FTIR. The peptides were incubated in the same buffer and temperature conditions. CD and FTIR results are presented in Figure 1.

![Figure 1 – Self-association kinetics of peptides STVIIE, GNNQQNY and ISFLIF. STVIIE and GNNQQNY CD spectra show a decrease in the ellipticity signal that may result from a rapid aggregation process. As ISFLIF could not be analyzed via CD due to low signal, FTIR was used instead, showing a maximum signal of the amide I band between 1625 – 1640 cm⁻¹, with a peak at 1636 or 1637 cm⁻¹, consistent with the typical of amyloid fibril β-sheet structure.](image-url)
Regarding the ISFLIF peptide, it was studied via FTIR spectroscopy since it had to be incubated at lower final concentrations (in order to dissolve at the desired physiological conditions of buffer and pH). As required, this peptide sequence presented a strong amide I maximum frequency of 1636/37 cm⁻¹, which is consistent with β-sheet structures [6].

Having established the peptides secondary structure content, AFM was employed in the subsequent studies, since it is a microscopic technique very useful for the study of amyloid fibrils because it allows the imaging of surfaces with high resolution and sensibility (Fig. 2).

The negative and positive controls behave as expected. STVIIE forms clearly the most promising amyloid fibril structure. It is also notorious that the peptide STVIIE is capable of forming amyloid fibrils in a shorter incubation time. After only 24 hours of incubation it already shows well-defined amyloid fibril morphology. In turn, the peptides GNNQQNY and ISFLIF only present amyloid fibrils near 2 weeks of incubation, with the former showing an atypical morphology. These findings fit perfectly well with the secondary structure analysis and support STVIIE, which solubilises well and possesses β-sheet structure, as a good candidate peptide, having been further studied regarding its morphology, as shown in Fig. 3. Briefly, the AFM-based morphological characterization of the STVIIE amyloid fibrils shows that these have about 80-100 nm in diameter and up to several nm in length, as suggested by previous similar approaches [1-5]. AFM, compared with other studies performed using electronic microscopy but at non-physiological buffer conditions (namely extremely low pH conditions), shows a very good match [5]. Thus, the STVIIE amyloid fibrils formed under these physiological buffer conditions of pH and temperature seem to be similar to standard amyloid fibrils constituting promising biomaterials. Furthermore, the use of AFM may allow the detailed characterization of the repeating patterns within the amyloid fibrils, namely in terms of size and shape. This may allow identifying the presence not only of mature fibrils but also of its repetitive constituent elements, the so-called protofibrils and/or oligomers. Since these have a role in toxicity and kinetics, these approaches may prompt our understanding of amyloidosis behaviour, with implications for toxicity and kinetics studies, besides our main goal of the development of an adequate amyloid-based nanomaterial compatible with biological applications.

![Figure 2](image1.png)  
**Figure 2** – AFM imaging of amyloid fibrils. As expected, negative controls did not form fibrils and the Aβ42 positive control forms typical amyloid fibrils [1-5]. As for STVIIE, GNNQQNY and ISFLIF, they all present amyloid-like fibril structures. Of these, STVIIE shows the most typical fibril morphology (consistent with the secondary structural content CD analysis). The data suggests a variety of amyloidogenic and aggregating behaviours in physiological conditions that can be functionalized further to tailored specific nanotechnology applications.

The negative and positive controls behave as expected. STVIIE forms clearly the most promising amyloid fibril structure. It is also notorious that the peptide STVIIE is capable of forming amyloid fibrils in a shorter incubation time. After only 24 hours of incubation it already shows well-defined amyloid fibril morphology. In turn, the peptides GNNQQNY and ISFLIF only present amyloid fibrils near 2 weeks of incubation, with the former showing an atypical morphology. These findings fit perfectly well with the secondary structure analysis and support STVIIE, which solubilises well and possesses β-sheet structure, as a good candidate peptide, having been further studied regarding its morphology, as shown in Fig. 3. Briefly, the AFM-based morphological characterization of the STVIIE amyloid fibrils shows that these have about 80-100 nm in diameter and up to several nm in length, as suggested by previous similar approaches [1-5]. AFM, compared with other studies performed using electronic microscopy but at non-physiological buffer conditions (namely extremely low pH conditions), shows a very good match [5]. Thus, the STVIIE amyloid fibrils formed under these physiological buffer conditions of pH and temperature seem to be similar to standard amyloid fibrils constituting promising biomaterials. Furthermore, the use of AFM may allow the detailed characterization of the repeating patterns within the amyloid fibrils, namely in terms of size and shape. This may allow identifying the presence not only of mature fibrils but also of its repetitive constituent elements, the so-called protofibrils and/or oligomers. Since these have a role in toxicity and kinetics, these approaches may prompt our understanding of amyloidosis behaviour, with implications for toxicity and kinetics studies, besides our main goal of the development of an adequate amyloid-based nanomaterial compatible with biological applications.

![Figure 3](image2.png)  
**Figure 3** – Morphological characterization of the amyloid fibrils formed by the peptide sequence STTVIE. Amyloid fibrils of STTVIE seem to repetitive patterns of specific sizes. The determination of these patterns and their study may shed light into the understanding of amyloidosis and the use of amylods in future nanotechnology applications.

**4 Conclusion**

Amyloid fibrils, which have been for long associated only with disease, became the focus of interest for the development of amyloid-based nanotechnology and nanomedicine approaches. Combining the data obtain from CD, FTIR and AFM, it is possible to conclude that among the three peptides tested, STTVIE is the most reproducible and amenable peptide for applications in nanotechnology as a novel biomaterial. The extension of this study into the determination of the repetitive patterns of STTVIE amyloid fibrils will further advance our understanding of amyloidosis in general and our ability to use this peptide sequence in nanotechnology.

**Acknowledgements**

This work was supported by Fundação para a Ciência e Tecnologia – Ministério da Educação e Ciência (FCT-MEC, Portugal), project PTDC/SAU-ENB/117013/2010. ICM acknowledges FCT-MEC post-doctoral fellowship SFRH/BPD/74287/2010.

**References**


Resumo
A pele apresenta-se como a primeira barreira física que o corpo humano dispõe para proteção, pelo que é importante que se encontre saudável. Para a sua avaliação e diagnóstico, a técnica de ultrassons apresenta-se vantajosa devido ao seu carácter não ionizante, não invasivo e acessível. Neste trabalho, foram desenvolvidas duas abordagens tendo em vista a caracterização da pele recorrendo a ultrassons. Para esse efeito, foram analisadas imagens ecográficas de pele de pacientes bem como imagens obtidas de fantomas. Uma das abordagens possibilitou a caracterização totalmente automática recorrendo a características texturais da imagem. A metodologia desenvolvida incluiu mais de 400 características texturais, 5 classificadores, 2 técnicas de seleção de características e um algoritmo de fusão de classificadores. A segunda abordagem permitiu a classificação de imagens de fantomas baseada na utilização de apenas três parâmetros acústicos: velocidade de propagação, coeficiente de atenuação e coeficiente de backscattering.

1 Introdução
A incidência de patologias da pele tem vindo a aumentar, justificada pela exposição excessiva ao sol, sem as devidas precauções. O diagnóstico precoce destas lesões aumenta a probabilidade de sucesso do seu tratamento. O método preferencialmente utilizado é a biópsia, no entanto, é uma técnica invasiva. Neste contexto, surge a técnica de ultrassons como não invasiva, não ionizante, inócuo e acessível, quando comparada com outras técnicas de imagionologia. No entanto, as imagens de ultrassons podem resultar em interpretações subjetivas, pelo que surge a necessidade de desenvolvimento de sistemas quantitativos capazes de auxiliar o diagnóstico. O principal objetivo do presente trabalho consiste no desenvolvimento de diferentes abordagens tendo em vista a caracterização de pele, recorrendo a técnicas de ultrassons.

2 Materiais

2.1 Fantomas
Foram desenvolvidos fantomas capazes de proporcionar propriedades acústicas similares às da pele. Os materiais selecionados para a sua criação foram a gelatina e microesferas de silicone. Várias concentrações destes dois materiais foram avaliadas, tendo-se concluído que a constituição que melhor simula a pele envolvia uma concentração de gelatina de 80g/l e uma concentração de microesferas de silicone de 40g/l e 60g/l, para a camada inferior e superior, respectivamente.

2.2 Imagens
Para o desenvolvimento das diferentes abordagens de caracterização recorreu-se à utilização de imagens de pele e imagens de fantomas. As imagens dos fantomas resultaram da conversão da intensidade dos sinais A-Scan adquiridos dos fantoma em 256 níveis de cinzento. As imagens de ultrassons da pele foram obtidas recorrendo a um ecógrafo portátil. Parâmetros como a frequência e a profundidade da imagem foram mantidos constantes com valores iguais a 13MHz e 4cm, respetivamente, permitindo um processamento mais simples. Ambas as imagens foram divididas em diferentes regiões de interesse (ROI), nomeadamente, 10, 20, 30, 40, 50 e 60. Foram avaliadas três profundidades: 1, 2.5 e 5 mm.

3 Métodos

3.1 Aquisição de sinais
A aquisição dos sinais foi realizada recorrendo a diferentes equipamentos incluindo, um transdutor de 25 MHz, um sistemas de eixos x-y-z, um emissor-receptor e um osciloscópio. Estes sinais foram armazenados usando o osciloscópio para posterior processamento e análise. O transdutor e o fantoma foram imersos em água, a qual atua como meio de acoplamento. A partir de cada fantoma foram extraídas 200 linhas A-Scan espaçadas de 100 µm entre si.

3.2 Parâmetros Acústicos
A avaliação dos três parâmetros acústicos relevantes (velocidade de propagação, coeficiente de atenuação e coeficiente de backscattering) foi realizada tendo em conta a abordagem que faz uso de um reflector plano, sendo considerado para tal o fundo do recipiente contendo o fantoma, para o caso do cálculo da velocidade e do coeficiente de atenuação e uma placa de alumínio, para o caso do cálculo do coeficiente de backscattering.

A velocidade de propagação foi obtida recorrendo à equação (1), onde \( t_1, t_2, t_r \) representam, respectivamente, o tempo de propagação do sinal de referência, do sinal da face anterior e o sinal da face posterior do fantoma e \( v_{agua} \) representa a velocidade de propagação na água.

\[
v_{fantoma} = \frac{v_{agua} (t_r - t_1)}{(t_r - t_2)},
\]

Para o cálculo do coeficiente de atenuação foi considerada a equação (2), onde \( F_s \) e \( F_p \) representam, respectivamente, as amplitudes espetrais do sinal do reflector e do sinal proveniente da interface fantoma-recipiente; \( \alpha_{agua} \) representa o coeficiente de atenuação na água e \( d \) representa a espessura do fantoma.

\[
\alpha_{fantoma} = \frac{20}{2d} \log \left( \frac{F_s}{F_p} \right) + \alpha_{agua},
\]

Para a avaliação do coeficiente de backscattering foi considerada a abordagem de Ueda et al [3], tendo em conta a equação (3).

\[
\eta(f) = \left( \frac{S_r(f, F_p)}{S_s(f, F_p)} \right) \times \left( \frac{R^2 k^2 a^2}{8 \pi d^2 \left[ 1 + \frac{k a^2}{4F} \right]} \right),
\]

onde \( S_r(f, F_p) \) e \( S_s(f, F_p) \) são respectivamente o espectro espectral do sinal do fantoma no foco do transdutor, \( S_s(f, F_p) \) é a amplitude espectral do sinal do fantoma proveniente do recipiente no foco do transdutor, \( R \) é o coeficiente de reflexão na fronteira água-recipiente, \( k \) é o número de onda (dado por \( 2\pi f/v \)), e \( a \) é o raio do transdutor, \( d \) é a largura da janela e \( F \) é a distância focal.

3.3 Classificação
Dos dois tipos de imagens (pele e fantomas) foram extraídas características texturais, derivadas da aplicação de estatísticas de primeira ordem, segunda ordem e ordem superior, Energia textural de Laws [4], Filtros de Gabor [5] e cálculo da atenuação em imagem [6]. Para o caso de imagens de fantomas foram também avaliados os três parâmetros acústicos.

As diferentes abordagens de classificação tiveram em consideração o uso de cinco classificadores de aprendizagem supervisionada: Redes Neurais Artificiais, Support Vector Machine, K-Nearest Neighbours, Naïve Bayes e Árvores de Decisão.
Foram implementados dois métodos de seleção de características e um algoritmo de combinação de classificadores, tendo em vista a melhoria do processo de classificação. Ambos os métodos de seleção de características se baseiam numa seleção sequencial, no entanto, o segundo método permite a reconsideração de características já rejeitadas e a remoção de características adicionadas [7]. O algoritmo de combinação de classificadores (Multiclassificador) recorreu a um voto por maioria dinâmica [8], tendo em conta a média de cinco parâmetros: sensibilidade, especificidade, precisão, exatidão e F-score.

A qualidade dos algoritmos de classificação foi avaliada a partir de cinco parâmetros de desempenho: sensibilidade, especificidade, precisão, exatidão e F-score. O F-score apresenta-se como uma medida que possibilita a avaliação de um desempenho global, sendo escolhido como parâmetro relevante para comparação das diferentes abordagens de classificação [9].

4 Resultados e Discussão

Tendo em vista a avaliação dos classificadores foram consideradas 34 imagens de quatro subgrupos: (1) imagens de pele sem lesão e (2) com lesão, (3) imagens de fantomas sem lesão e (4) com lesão. Foram considerados os desempenhos obtidos para três grupos de imagens: pele, fantomas e o conjunto de imagens de pele e de fantomas. Foram usados os dois métodos de seleção de características e uma combinação das suas saídas, definindo-se um limite máximo de características igual a 20. A divisão dos dados em teste e treino foi realizada recorrendo à técnica de k-fold cross validation, considerando um valor de k igual a 10.

4.1 Imagens da Pele

A análise de classificação de imagens da pele resultou na seleção do classificador Redes Neuronais, 50 ROIs por imagem com uma profundidade de 2.5 mm. Verificou-se que a combinação dos dois métodos de seleção de características, apesar de possibilitar uma redução do custo operacional, resulta numa diminuição do desempenho (F-score igual a 77.0%, quando comparado com os 89.9% alcançados recorrendo ao segundo método de seleção de características).

4.2 Imagens de Fantomas

A análise de imagens de fantomas foi realizada considerando-se três grupos de características: (1) texturais (2) acústicas e (3) combinação das mesmas. Para as imagens de fantomas analisadas observou-se que 50 ROIs por imagem com uma profundidade igual a 2.5 mm, foi a combinação que resultou em melhores desempenhos. Observou-se que a combinação de características texturais com parâmetros acústicos proporciona um maior F-score (89.7%) quando comparado com os resultados obtidos considerando apenas as características texturais (88.1%) ou os parâmetros acústicos (62.8%). As Redes Neuronais proporcionaram os melhores resultados exceto, considerando apenas parâmetros acústicos, onde o Multiclassificador se apresentou vantajoso. Isto indica que na presença de menor quantidade de dados, devem ser considerados diferentes classificadores para obtenção de melhores resultados.

4.3 Imagens da Pele e de Fantomas combinadas

Para este caso observou-se que os melhores resultados ocorrem para 50 ROIs por imagem com profundidade de 2.5 mm e Redes Neuronais, com um F-score igual a 77.2%. Comparando estes resultados com os obtidos para imagens da pele verificou-se que, apesar de ocorrer uma diminuição do F-score, ocorre um aumento da sensibilidade (de 83.7% para 86.7%). Observou-se que a diminuição do valor de F-score deve-se a um aumento dos falsos positivos. Vários autores sugerem que a sensibilidade apresenta uma maior importância, pois é mais grave a ocorrência de falsos negativos do que de falsos positivos. A combinação de imagens da pele e de fantomas apresenta-se assim vantajosa na correta identificação de casos positivos.

4.4 Melhoria do desempenho

Utilizando 50% de imagens com lesão e 50% de imagens sem lesão resulta num elevado número de ROIs sem lesão. A fim de melhorar o desempenho foi considerado um igual número de ROIs com e sem lesão. Esta consideração resultou num equilíbrio entre verdadeiros negativos e verdadeiros positivos e entre falsos negativos e falsos positivos, proporcionando valores de desempenho superiores, tal como se mostra no gráfico da figura 1.

Figura 1: F-score das diferentes abordagens de classificação, considerando igual número de imagens e de ROIs com e sem lesão.

5 Conclusões

O presente trabalho mostrou que é possível a caracterização de imagens da pele recorrendo a duas abordagens: classificação recorrendo a características texturais da imagem assim como recorrendo a parâmetros acústicos. Concluímos também que a consideração de um igual número de dados de entrada com e sem lesão possibilita melhores valores de desempenho, pelo que será possível um aumento de desempenho na presença de um maior número de imagens. Os parâmetros acústicos revelaram-se uma opção viável para o diagnóstico, considerados separadamente ou em combinação com características texturais.

Referências

Abstract

The autonomous use of tools by robots is a topic where studies are still rare even though much attention is given to the affordances of objects. In this paper we propose an extension to the learning object affordances model presented by Montesano et al. in [1] by including tools as an additional node into the Bayesian Network of affordances. We also present an automatic method for gathering the data required in the learning of such a network using the iCub simulator.

1 Introduction

The use of tools by an animal species is considered a major sign of intelligence, as it shows the ability to understand relations between object features and the resulting effects of interaction with the environment. Their use is of the utmost importance to humans as they are used in most of human activities. For a humanoid robot to be able to perform the same everyday tasks humans do, tool manipulation is imperative.

In this paper, similarly to Stoytchev [2] and Jain et al. [3], we explore the extension of reach that a tool can provide to alter the position of another object, which is based on the definition of tool given by Beck [4]: “Tool use is the external employment of an unattached environmental object to alter more efficiently the form, position, or condition of another object, another organism, or the user itself when the user holds or carries the tool during or just prior to use and is responsible for the proper and effective orientation of the tool.”

Closely associated with the concept of tool is the concept of affordances introduced by Gibson [5], which describes what an organism directly perceives of an object as doable (affordable actions) to said object. The understanding of such properties and relations can lead to emergence of planning behaviors and more complex activities.

In order to achieve a method which enables a robot to learn how to successfully use tools in the best way possible, this paper introduces a way to represent tool affordances, grounded to the individual experience of the robotic agent, and also an autonomous behavior for collecting the data required for learning such representation. However, deriving analytical or physical models of the effect that a tool produces on objects is hard due to uncertainty in contact interactions, friction, impact forces and sensor noise. Therefore we adopt a learning by exploration methodology to obtain causal models of tool affordances.

2 Tool Affordances Modeling

The affordances of a tool are tightly related not only with the tool itself but also with the presented environment (objects), the individual motor capabilities of an agent and the effects on the environment recognizable by said agent. In order to successfully use tools, there is a need to encode the relation between these 4 elements: Tool, Object (Environment), Action, and Effects.

In [1], Montesano et al. introduced a way of encoding the dependencies between actions, directly applied to objects and its resulting effects as a Bayesian Network. In this paper we propose the introduction of an additional node to the network which encodes the use of tools as shown in Figure 1.

![Proposed Bayesian Network model of tool affordances](image)

Figure 1: Proposed Bayesian Network model of tool affordances

This Bayesian Network graphical model of affordances expresses the joint probability of an event tuple $P(T=\text{tool}, O=\text{object}, A=\text{action}, E=\text{effect})$ as a factorization: $P(T,O,A,E) = P(E|T,O,A)P(T).P(O).P(A)$. With this model it is possible to use the derived affordances to: predict the effects on a given object using a specific tool $P(E|T,O,A)$; plan the best action which produces the desired effects on an object given a tool $P(A|T,O,E)$; select the most likely object to produce the desired effects given a tool and action $P(O|T,A,E)$; or select the best tool given an object, action and desired effect $P(T|O,A,E)$.

![Proposed scenario, iCub humanoid robot with a tool and an actuated object](image)

Figure 2: The proposed scenario, iCub humanoid robot with a tool and an actuated object

2.1 Visual Descriptors

A set of visual descriptors is used for tool and object recognition and classification. These describe shape, size, position and color of the segmented objects and tools.

The descriptors, calculated in the 2D camera image plane, are as follows:

- Normalized x and y coordinates of the center of the enclosing rectangle.
- Normalized width of the enclosing rectangle.
- Normalized height of the enclosing rectangle.
- Angle (orientation) of the enclosing rectangle.
- Hue normalized color histogram of the pixels inside the object’s region.
- Area (number of pixels).
- Convexity - ratio between the perimeter of the object's convex hull and the perimeter of the object's contour.
- Eccentricity - ratio between the minor and major axis of the minimum-area enclosing rectangle.
- Compactness - ratio between the area object and its squared perimeter.
- Circleness - ratio between the object area and the area of its enclosing circle.
- Squareness - ratio between the object area and the area of its minimum-area enclosing rectangle.

By continuously tracking an object during an interaction with it and storing both the visual description before and after the interaction, the effects can be recognized and classified as change in position, color and/or shape of the object.
Additionally, we propose to divide a tool in 3 different segments, each with its own visual descriptors parameters:

- **Handle**, which is the graspable part of the tool.
- **End-effector**, the end extremity of the tool which will come into contact with the object.
- **Body**, the connecting part between the handle and the end-effector.

With this separated parameterization of each part of a tool we hope to give a better generalization capability to the learning of the affordances, as for example certain tools might not afford certain actions not because its end-effector is inappropriate but because the tool might be too short to reach the target object or simply not graspable by the robot.

3 Simulated Learning Environment

In order to learn the structure and parameters of the affordances Bayesian Network for a specific set of tools, objects, actions and effects, an experimental setup using the iCub simulator was developed and automated.

This simulated setup enables the iCub robot to repeatedly perform automated trials on a provided environment at a much higher rate than using the real robot, which will be posteriorly used to validate the results. Such a trial involves 5 steps:

1. **Observation of the tool**. A specific tool is presented to the robot and its visual descriptors are calculated and stored. Shown in Figure 3, Figure 4 (A) and (B).

2. **Grasping of the tool**. The robot grasps the tool by the handle after it is automatically placed on his hand.

3. **Observation of the object**. A specific object is presented to the robot; its visual descriptors are calculated, stored and starts to be actively tracked. An example of a segmented object during action with tool (D).

4. **Act**. The robot is commanded to perform an action on the object using the tool. The action and its parameters are stored. Figure 4 (D) and Figure 5 show an example of an action being performed by the robot.

5. **Re-observation of the tracked object**. The visual descriptors of the object are recalculated after the interaction with the tool and its parameters are stored. Initially the set of available actions, tools and objects will be limited to the simple case of:

- One parameterized action: \( \text{Push}(x, y, z, \theta, \text{radius}) \), where \( x, y \) and \( z \) are the 3D coordinates where to act with the tool end-effector, \( \theta \) the angle of approach, and \( \text{radius} \) the radial offset from which the tool end-effector will approach the \( x, y, z \) position.
- Three uniformly colored tools: Stick, L-stick and T-stick.
- Three uniformly colored objects: Box, Ball and Cylinder

These sets will later be expanded to include more complex and varied behaviors and objects.

4 Future Work

This paper gives an overview of a Bayesian Network model for tool affordances and an experimental setup for the iCub robot which enables the collection of data required to generate such network.

Although the outline of the experimental setup is established there is still work to be done in both the motor and visual elements, as additional actions and visual descriptors are being evaluated. More importantly multiple trials combining all the available tools, actions, and objects are still needed in order to generate the data that will be used in the continuation of this work which is to learn the actual structure and values of the affordances Bayesian Network.

5 Acknowledgements

This work was partially funded by the EU Project POETICON++ and FCT [PEst-OE/EEI/LA0009/2013].

References


Abstract

In this paper we propose an automatic system for the visual-tracking of buses in a parking lot, by using a set of Pan-Tilt-Zoom (PTZ) cameras. It is assumed that the parking lot has specific entry points, so that the places from where the buses come in are known beforehand. To detect the buses a background subtraction method is used, being then made an initial estimation of the bus position through backprojection. This estimate is then refined by an algorithm that also provides the bus orientation. The final estimate is then used in a EKF (Extended Kalman Filter) filter to provide an estimation of the bus next position, which allows to decide the camera movement. To test the system a simple simulator was developed using Matlab and Virtual Reality Modelling Language (VRML).

1 Introduction

Task automation has been increasing in a variety of industries and services, due to its capability of improving either efficiency and effectiveness. In this context the work presented in this paper is proposed as a method to automatically track buses moving in a parking lot. The system consists of a set of multiple PTZ cameras placed at certain locations of the parking lot, that begin to track and detect the buses from the moment they enter it.

In [7] a number of motion modalities for one pan-tilt camera are assessed with respect to omni-awareness or, more precisely, maximizing the percentage of events found. Starzyk and Qureshi [6] consider multiple PTZ cameras to track pedestrians (moving events). They design a behavior based architecture which handover tracking-inter-camera and maximize the zoom of each camera while not loosing tracking of all pedestrians. In our work the events to find and track are moving buses. Contrarily to the motion of people, the motion of buses can be predicted. In this work we take the predictions of the motion into account in the design of PTZ control.

2 Camera Model and Bus Detection

The pin-hole camera model [4, 5] is used in this work to represent the relationship between world, $M$, and image, $m$, coordinates:

$$m \approx PM = K[R | t] M$$

where $P$ is called the projection matrix, $K$ and $R$ the intrinsic and extrinsic parameters matrices respectively and $t$ is the translation vector.

It is assumed that when a bus enters the parking lot a signal is sent to notify the system, which then moves the field of view of an available camera to the adequate entry point. From this time onwards the bus detection is made through a background subtraction algorithm, and a measure of the bus position is generated by using backprojection. Then an optimization algorithm uses a bus 3D model to improve this estimation and to calculate the bus orientation.

2.1 Background Subtraction and Bus Pose Estimation

In order to detect a bus, we start by building a model of the background for each of the PTZ cameras [3]. This background model allows acquiring a background image at any pan-tilt-zoom configuration. The current (real) image is then subtracted to the background image, resulting a logical mask whose pixels indicate the bus pixels.

The center of mass of the detected bus pixels allows estimating coarsely the bus location. Backprojection [4] is applied to the mass center (pixel location) subject to setting the Z coordinate to be in the ground plane. In practice this corresponds to solving a matrix equation, $M = C + \alpha D$, where $M$ is a point in world coordinates (as in equation 1), $C$ the projective center, $D$ is a point in infinity and $\alpha$ is a scaling factor.

The Kalman filter is a linear quadratic estimation method to compute estimates of unknown variables, from measurements corrupted with noise.

2.2 Fine Tuning the Pose Estimation

To fine tune the (coarsely) estimated pose it is used a minimization algorithm that finds the local minimum of a cost function of several variables:

$$\hat{X} = \arg \min_{X, Y, \theta} F(X, Y, \theta)$$

where $A$ and $B$ are both binary masks (see figure 1 for an example). In the case of $A$ it is obtained from the background subtraction algorithm of section 2.1. Relatively to $B$ it is a synthetic mask, generated by placing a 3D model of the bus in an image of the background and then using it in the background subtraction algorithm mentioned in 2.1. Initially the 3D model is placed at the position $X, Y$ of the world referential frame, computed from the backprojection mentioned in 2.1 and with the orientation $\theta$ obtained from the EKF (see section 3.2) prediction step. With each iteration of the minimization algorithm, a Levenberg-Marquardt like algorithm, the $X, Y, \theta$ values are updated so that the value of the cost function $F(X, Y, \theta)$ approaches zero. When the search ends, the masks $A$ and $B$ will ideally be overlapping perfectly.

$$F(X, Y, \theta) = \min \frac{|A \cap B|}{|B|}$$

3 Bus Tracking Algorithm

For each bus being tracked there is an associated EKF which is used to predict the bus position in the next sampling time. The measurements correspond to the estimation obtained from the bus detection algorithm mentioned in section 2. The camera pan and tilt angles are then adjusted in order to position the center of view at the EKF estimation. This guarantees that in the next sampling time the bus will be kept visible and close to the image center.

3.1 Bus kinematics model

The bus kinematics model used assumes that: (i) the front wheels can spin around their axis but have no traction, (ii) the back wheels have traction but do not spin or slide, (iii) the body and wheels are assumed to be rigid bodies [1, 2]. From these assumptions and the geometric relations shown in figure 2(a) it was then generated the following discrete kinematics model, that is used by the Extended Kalman filter in the prediction step:

$$\begin{align*}
x_{k+1} &= x_k + T \cdot v_k \cos(\theta_k + \Phi_k + \frac{\pi}{2} \sin(\Phi_k)) \\
y_{k+1} &= y_k + T \cdot v_k \sin(\theta_k + \Phi_k + \frac{\pi}{2} \sin(\Phi_k)) \\
\theta_{k+1} &= \theta_k + T \cdot \frac{\pi}{2} \sin(\Phi_k) \\
v_{k+1} &= v_k \\
\Phi_{k+1} &= \Phi_k
\end{align*}$$

The state variables are the position $x, y$ of the bus front axle shaft center $(x_f, y_f)$ and the bus body orientation angle $\theta$. The model inputs are the bus linear velocity $v$ and the front wheels angle $\Phi$ relatively to $\theta$, both assumed to be constant. In what concerns the constants $T$ and $L$ they are the sampling interval and the front wheel distance respectively.

3.2 Extended Kalman Filter

The Kalman filter is a linear quadratic estimation method to compute estimates of unknown variables, from measurements corrupted with noise.
mean Gaussian noise. As the original Kalman filter could only be used on linear systems the Extended Kalman filter was used instead. It assumes a system of the form:

\[
\begin{aligned}
X_{k+1} &= f(X_k, \xi_k) \\
\eta_k &= h(X_{k+1}, \eta_k)
\end{aligned}
\]  

The variables \(X_k\) and \(\eta_k\) are the state variables and the measurements respectively. In what concerns \(\xi_k\) and \(\eta_k\) they are zero mean multivariate system and observation Gaussian noises. The functions \(f\) and \(h\) correspond to non-linear functions which are linearized. Then the Extended Kalman filter equations are applied to obtain the \textit{a priori} and the \textit{a posteriori} estimations.

4 Complete System

The complete system takes into account three main components: buses, cameras, and the 3D world. Buses have autonomous motions (are driven by on board drivers). The PTZ cameras are mounted at known positions and orientations of the 3D world. PTZ cameras are controlled automatically to detect and track the buses. The 3D world, the parking-lot floor, is described as a plane.

At every second of time elapsed, the following tasks are run: (i) Predict the buses position and orientation with the EKF and compute corresponding uncertainty ellipses. (ii) Assign a bus to a camera taking into consideration a FIFO multitasking algorithm which takes into account the uncertainty ellipses’ areas and distances of the cameras to the buses in pan, tilt and zoom units. (iii) Move each camera to image the coordinate predicted for its assigned bus \((X, Y)\). (iv) For every image acquired by the cameras, observe the buses position using the algorithm described in section 2. (v) Update the EKF filters using the observations of the buses.

5 Experimental Results

In order to test the complete system a 3D simulator built was developed in MATLAB using VRML. The buses have predefined trajectories. The PTZ cameras are controlled automatically.

The simulated system encompasses two cameras \((C_1\) and \(C_2)\) and four buses \((B_1, B_2, B_3, B_4)\). Figure 2(b) shows an aerial view of the parking lot at a simulation iteration where the four buses were all in the scene. The center of the world referential frame is at the center of the floor plane of figure 2(b). The cameras were positioned at the red dots location and oriented in the direction of the “red triangles” base associated to each red dot. Typical images acquired by the camera are shown in figures 2(c) and 2(d).

At the end of the experiment the plots presented in figure 3 were obtained. In these graphs the rectangles position and orientation represent the estimates obtained from the EKF update step for each bus, at a given simulation moment. The elliptical lines are the limits of the uncertainty area associated with the EKF estimation. As can be noted in these graphs the program was able to keep track of the positions of the buses during all the simulation. There were however certain positions where the estimation was not accurate. However the system was always able to recover from the errors and continue tracking the buses.

6 Conclusion and Future Work

In this paper an algorithm to detect and track multiple buses was developed as well as a simulator to test it. The approach used showed encouraging simulation results.

There are however some improvements that need to be applied to make this project feasible in reality. For example the background subtraction algorithm has to be improved so that it can handle the lighting changes that would occur in the real world. Another example, a camera controller needs to be developed to deal with the dynamics of the motors moving the camera.

The simulator itself also has much space for improvement. As an example, a number of light changes that occur in real world can be simulated by an artificial sun moving according to the time of the day.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE/EEI/LA0009/2013, by the FCT project PTDC/EEACRO/105413/2008 DCCAL, and by the FCT project EXPL/EEI-AUT/1560/2013 ACDC.

References

Object tracking with UAVs

João Palma 1,2
Arnaldo Abrantes 1
Pedro Mendes Jorge 1

1 Department of Electronics and Telecommunications and Computer, ISEL
2 TEKEVER Information Technology

Abstract
This paper presents an automatic object tracking system that processes video captured by a camera installed in a UAV. The system returns the position of a given target which can be integrated with the GPS navigation control system of the aerial vehicle.

1 Introduction
The use of UAVs is growing up at an impressive speed. Every day we heard about military operations concerning UAVs, not only for vigilance purposes, but also for operations involving the attack of predefined targets. [4]

A large number of projects have been presented in the last few years, related with tracking using UAVs. COCOA [2] and MODAT [3] are two examples of systems that perform real time tracking of objects captured by a video camera installed on UAVs. The implementation of both of these systems adopts an architecture where the output (estimation of the target position) is obtained by applying to the input (captured video) a chain of three consecutive standard image analysis techniques: Image Registration, Image Segmentation and Object Tracking. According to MODAT, this system has a success rate of 80%. However, the authors refer also a false positive rate of 18%.

In this work, it was developed an automatic object tracking system that processes at real time the video captured by a camera installed at UAV. The system returns the position of the desired target (an object that is moving on the field of view of the camera) and can be integrated with the GPS navigation control system of the aerial vehicle.

The proposed system is based on the Mean shift algorithm to perform object tracking. In order to improve the results, this standard tracking technique was enhanced by previously filtering a back projected image (used as the input of Mean shift), using the target predicted position given by a Kalman filter.

2 Object tracking with UAVs
The proposed system is described by the block diagram presented in Figure 1. The system contains three main modules, which are summarized in the following subsections.

![System architecture with its three modules: Image Registration, Motion Detection and Object Tracking](image)

Figure 1: System architecture with its three modules: Image Registration, Motion Detection and Object Tracking

2.1 Image Registration
This first module is responsible to align the captured frames by incrementally making a mosaicing, where all the frames belong then to a unique world referential. This is a crucial operation to make possible to detect object motion with image subtraction. Image registration is accomplished by combining the following techniques: camera calibration, image homography, feature extraction, and optical flow estimation.

A camera calibration method [1] is used to correct geometric deformations present in the video frames. In this kind of projects, these deformations are usually strong, as the used lenses have a large field of view, which generate significant radial and tangential distortions (Fish-Eye Lens). The align phase uses feature extraction to detect interest points (corners) and optical flow to follow those features in the next frame. The set of corresponding points is then used to estimate a homography matrix, which is finally used to align consecutive frames into the same geometric referential.

2.2 Motion Detection
Motion detection uses two different techniques: image subtraction and color information.

In the first one, the aligned frames are returned from the previous module and subtracted to obtain a difference image that can be used as a starting point to detect the moving objects. A second approach is the use of a color histogram to model the desired target, and subsequently compute, for each new frame, the probability of each pixel color was produced by the moving target. This new image, containing the referred probabilities, is usually designed as back projected image.

2.3 Object Tracking
The system tracks one of the objects returned by the motion detection module: that one selected by the user.

Tracking is then accomplished by combining Meanshift and Kalman filter algorithms.

Meanshift is used to estimate the object location based on the local maximums (modes) of the back projected image (i.e., a probability density function). The algorithm performs an iterative local optimization, searching in each iteration for the maximum of the pdf in a small square search window, centered at the previous estimation. Kalman Filter is an algorithm that produces estimates of state variables based on a series of measurements (observations). In this model, it is assumed that prediction and measurement values are affected by white Gaussian noise with zero mean. A dynamic model with constant velocity is used in this work.

The tracking algorithm operates as following: first, the target’s location is predicted using the Kalman filter prediction equation, then Meanshift is used to refine the predicted position, which is finally used as the observation in the last step of the Kalman filter.

2.3.1 Filtering back projected image
In several tracking problems, there are many small potential targets (each one with only a few numbers of pixels) and most of them indistinguishable in terms of color (see Figure 2.a). In this case, color information (e.g. color histograms) is insufficient to detect the target as can be seen by observing the high noisy back projected image (Figure 2.b). Figure 2.a is an example of the back projected image from a window near the target location of this sequence. All the vehicles in this sequence exhibit the same color and, based only in this information, it is very difficult to estimate the correct target location.

![Binary representation of back projected image (a), based on the color histogram of car A (a)](image)

Figure 2: Binary representation of back projected image (a), based on the color histogram of car A (a)

The existence of several local maximum within the Meanshift search window can become a point of failure when applied to object tracking. To alleviate this difficulty a Gaussian filter, centered on the region identified as the predicted target location, is applied to the back projected image. This filtering operation is able to reduce the noise in the back projected image and therefore improving the results of Meanshift (Figure 3.b).
Meanshift results (Figure 6.a) the closest regions will have a much lower probability, increasing the Gaussian’s center is fixed at Kalman’s predicted point (Figure 5.a), Gaussian filter to the previous back projected image (Figure 5.b). Once generates a new back projected image (Figure 6.b) by applying a approach to reduce the processing errors of Meanshift. This technique can lose the object of interest. Correcting the back projected image is an methodology of this work.

Figure 2.a. Gaussian is centered at point A, predicted by the Kalman

If occurs an error in object detection caused by Meanshift, Kalman filter tends to fail whenever two similar (in color) targets are close, as in this case, Kalman filter may be driven by a false observation. In order to alleviate this difficulty and enhancing the robustness of the method, the back projected image is previously processed, filtering out the data further from the predicted position of the target, before being used by Meanshift algorithm. The results obtained with this systems show that it is capable of accurately tracking a small target even if other distracting targets are close. However, there are situations when the system loses its target and in that case it cannot recover from this failure. This is the main problem of the proposed method which can be ameliorated by extending it to a multimodal technique like, e.g., particle filters. The target trajectory detected with this algorithm is still not used to control the UAV. The integration with the GPS navigation control system will be a next stage of this project. Some future work on multi-target tracking should also be addressed, namely, to maintain multiple targets in the UAV field of view.

Figure 9: Tracking results with back projection matrix filtered

Figure 8: Tracking results without filtering back projection matrix

4 Conclusion and Future Work

This paper describes a real time tracking system for UAVs. The proposed method is based on the combination of Meanshift and Kalman algorithms, using as input back projection matrix obtained by measuring, at each frame, the difference between local color histogram and a model histogram representing the tracked target. The systems tends to fail whenever two similar (in color) targets are close, as in this case, Kalman filter may be driven by a false observation. In order to alleviate this difficulty and enhancing the robustness of the method, the back projected image is previously processed, filtering out the data further from the predicted position of the target, before being used by Meanshift algorithm. The results obtained with this systems show that it is capable of accurately tracking a small target even if other distracting targets are close. However, there are situations when the system loses its target and in that case it cannot recover from this failure. This is the main problem of the proposed method which can be ameliorated by extending it to a multimodal technique like, e.g., particle filters. The target trajectory detected with this algorithm is still not used to control the UAV. The integration with the GPS navigation control system will be a next stage of this project. Some future work on multi-target tracking should also be addressed, namely, to maintain multiple targets in the UAV field of view.

References


Minha moralidade de referência na detecção, caracterização, acompanhamento do grau de atividade da doença e como guia de biópsia pulmonar da grande maioria das doenças pulmonares. O elevado volume de imagens geradas pelos atuais scanners e a complexidade da análise de imagens de TCAR na presença de padrões pulmonares patológicos tem incentivado o desenvolvimento de métodos computorizados de análise de imagem.

Uma das doenças pulmonares mais frequentes é o enfisema, uma doença obstrutiva crónica, que em fases avançadas afecta severamente a qualidade de vida dos pacientes. A sua detecção precoce é essencial para a identificação do factor de risco que contribui para a progressão da doença, dando a oportunidade ao paciente para mudar as suas rotinas, se for caso disso.

Em imagens TCAR o enfisema surge como regiões de baixa atenuação, portanto a zonas mais escaras [1]. Muitos métodos baseiam-se neste facto para a detecção e quantificação de enfisema [2]. No entanto, em fases iniciais da doença ou em situações em que as zonas de baixa atenuação de encontram dispersas pelo parénquima normal, estes métodos mostram-se bastante insensíveis. Na Figura 1 podemos observar o aspecto visual dos padrões pulmonares em imagens de pulmão saudável e na presença de diferentes tipos e fases de enfisema.

O objectivo deste estudo consiste em estudar a textura de padrões pulmonares, em imagens TCAR, através da análise da sua lacunaridade, de forma a permitir a diferenciação entre padrão pulmonar normal e enfisema. A lacunaridade é uma propriedade fractal que descreve a textura de um fractal através da análise da distribuição do tamanho das lacunas do fractal [3, 4].


Ao longo deste resumo é efectuada uma introdução ao assunto abordado e técnicas utilizadas. É descrito o algoritmo usado no cálculo da lacunaridade diferencial. É descrita a base de dados utilizada, a metodologia adoptada e resultados obtidos. Por fim, são extraídas conclusões.

Figura 1: Imagens de TCAR do tórax de pulmão normal (A) e de pulmão na presença de enfisema (B a D).

1 Introdução

Atualmente, a TCAR do tórax é considerada a modalidade de referência na detecção, caracterização, acompanhamento do grau de actividade da doença e como guia de biópsia pulmonar da grande maioria das doenças pulmonares. O elevado volume de imagens geradas pelos atuais scanners e a complexidade da análise de imagens de TCAR na presença de padrões pulmonares patológicos tem incentivado o desenvolvimento de métodos computorizados de análise de imagem.

2 Lacunaridade Diferencial

A lacunaridade é uma medida complementar da DF. Enquanto a DF quantifica o espaço preenchido pelo fractal, a lacunaridade descreve como esse espaço é preenchido. Um valor baixo de lacunaridade corresponde a padrões espaciais homogéneos e valores alto de lacunaridade a padrões heterogéneos. O algoritmo proposto por Dong [8] para calcular a lacunaridade diferencial é apropriado para imagens de níveis de cinzento, muito comuns na grande maioria das aplicações de imagens em situações reais. Este algoritmo baseia-se nos algoritmos Gliding Box Algorithm e Differential Box Counting [9, 10].

Para o cálculo da lacunaridade diferencial a imagem é dividida em janelas sobrepostas de tamanho \( w \times w \) que se deslocam ao longo de toda a imagem. Detecta-se a caixa de \( r \times r \) pixels. O tamanho da janela \( w \) deve ser ímpar e \( r < w \). De acordo com a gama de variação dos níveis de cinzento correspondentes aos pixels da caixa é construída uma coluna de caixas sobrepostas de dimensão \( r \times r \times r \). A cada caixa que constitui a coluna é atribuído um número consecutivamente de baixo para cima da coluna (1, 2, ...). Considerando que o valor dos pixels mínimo e máximo residem nas caixas \( u \) e \( v \), respectivamente, a altura relativa da coluna na posição \( (i, j) \) é dada por:

\[
\begin{align*}
\hat{n}(i, j) &= v - u - 1. \\
M &= \sum_{i, j} n(i, j). 
\end{align*}
\]

Uma vez que a caixa percorre toda a janela, a massa da janela \( w \), calculada numa determinada posição é dada por:

\[
M = \sum_{i, j} n(i, j).
\]

Seja \( n(M, r) \) o número de janelas \( w \) com massa \( M \), calculada através de uma caixa de \( r \) pixels. A função de probabilidade \( Q(M, r) \) obtém-se dividindo \( n(M, r) \) pelo número total de janelas. A lacunaridade diferencial da imagem, à escala \( r \), dada uma janela \( w \), é definida como
3 Base de Dados

O conjunto de imagens usado nos testes experimentais foi organizado em colaboração com o Departamento de Radiologia do Hospital da Universidade de Coimbra. O médico radiologista analisou as imagens e contorna as regiões que contêm padrões pulmonares típicos. Cada região é caracterizada de acordo com o padrão presente, dados do paciente e da secção de onde foram extraídas. Os dados recolhidos podem ser usados no treino de sistemas de diagnóstico assistido por computador baseados em classificadores supervisionados, uma vez que cada região de interesse se encontra catalogada de acordo com o padrão presente.

As imagens de TCAR foram adquiridas através de um scanner LightSpeed VCT 64 da General Electric Healthcare. Os exames de cada paciente são guardados em imagens individuais em formato DICOM (Digital Imaging and Communications in Medicine), correspondentes às secções obtidas. As regiões de interesse foram extraídas de 290 imagens de 82 pacientes (55 homens e 27 mulheres) com idade média de 65±15 anos. As imagens que integram este estudo possuem para regiões a parênquima normal e a enfisema. As regiões foram extraídas de diferentes zonas do parênquima pulmonar, de diferentes tipos de enfisema e com graus de severidade diferentes. Cada uma das regiões selecionada pelo médico é dividida em regiões mais pequenas de 40 x 40 pixels [11,12].

4 Metodologia e Resultados Obtidos

O parênquima pulmonar saudável ou na presença de patologias possui uma textura muito rica. A sua análise é crucial para a elaboração de um diagnóstico correto, quantificação da extensão da doença e sua monitorização. O aspecto visual de imagens pulmonares depender de um elevado número de factores. Mesmo o padrão considerado normal pode ter aspectos visual muito diferentes consoante a zona do pulmão de onde é extraída a região de análise.

Um aspecto muito importante em análise de textura pulmonar é a escala considerada. Uma textura pode ser homogénea quando analisada numa escala pequena e heterogénea quando analisada em escalas maiores, e vice-versa. Assim, é importante recorrer a técnicas multiescala. A lacunaridade diferencial pode ser calculada para todas as combinações de \( r \) e \( w \). O valor de \( r \) pode ser mantido constante à medida que se varia o valor de \( w \). Analisando os resultados obtidos, podemos verificar quais as combinações que melhor permitem diferenciar os padrões pulmonares em estudo.

A lacunaridade diferencial foi calculada para várias combinações de \( r \) e \( w \) (respeitando a condição \( r < w \)) para as regiões que constituem a base de dados de estudo. Após testes efectuados para diferentes gamas de \( (r,w) \) verificámos que os valores da lacunaridade diferencial se mantêm praticamente constantes para valores de \( w \) superiores a 25% do tamanho da região, cerca de 10 pixels.

Para tamanhos de janelas com \( w < 10 \) os valores médios obtidos para a lacunaridade diferencial em regiões de enfisema e de pulmão normal são significativamente diferente. Das gamas de variação dos parâmetros consideradas verificámos que as curvas da lacunaridade média calculadas para os dois tipos de padrões em estudo são mais discriminatórias para valores pequenos do tamanho da janela. A Figura 2 mostra as curvas obtidas para um valor constante de \( r=2 \), fazendo variar \( w \) de 3 a 10. Observando estas curvas, verifica-se que os valores de lacunaridade obtidos permitem diferenciar os padrões pulmonares em estudo.

A importância do número de níveis de cinzento das regiões foi também objecto de estudo. Foram calculadas curvas de lacunaridade diferencial média para 32, 64, 256 níveis de cinzento e para escala original de Hounsfield. Os resultados obtidos mostram que as curvas mais discriminatórias correspondem à quantização das regiões na escala de Hounsfield, o que mostra a sensibilidade do algoritmo escolhido à variação dos níveis de cinzento no cálculo da lacunaridade diferencial.

5 Conclusões

A análise de textura em imagens de TCAR tem um papel determinante na correcta caracterização, quantificação e acompanhamento do enfisema, uma doença obstrutiva crónica. Neste estudo foi avaliado o potencial da lacunaridade diferencial na discriminação de padrões pulmonares normal de enfisema. As curvas obtidas para algumas combinações de \( (r,w) \) mostram o potencial da abordagem escolhida.

Referências

Abstract

Personal, inexpensive, easy-to-use, teleoperated, autonomous, mobile robots are starting to be a reality and are expected to be widespread within some few years. Fostering the development of these robots implies testing various hardware and software technologies. In this work we propose developing the communications with the robot through an IP wireless camera. The rational is that the required bandwidth for motion commanding and reporting is much lesser than the one required for video streaming. Therefore, robot commanding can be embedded non-disruptively within the video streaming. In this paper we detail the base components forming the robot, show navigation experiments based in MonoSLAM, and propose a methodology to regulate the accumulation odometry error associated with map-less MonoSLAM.

1 Introduction

Current home personal robots are starting to have affordable prices. These robots can be vacuum cleaners such as the iRobot Roomba, can be telepresence robots, as the Anybots’ QA, or can be simply Mobile Webcams, such as the WowWee’s Rovio. Most of these robots have in common the combination of mobile robotics, video cameras and wireless communications. Communications are in many aspects the bottleneck of the robots.

In this work we propose using the wireless network of surveillance cameras as a basis to build networked mobile robots. The objectives of this work are threefold: (i) assembling one Axis 207w camera and one Arduino Uno; (ii) developing software to make possible the communication between the user’s pc, the Axis camera and the Arduino; and (iii) Designing a homing strategy which enables the robot to return autonomously to its initial position.

2 Hardware and Software Setup

A car is assembled based in a network (IP) camera, one Arduino Uno, one Motor Shield and two DC motors for steering and propulsion (see fig. 1(a)). A Java GUI from [2] is used to control the car in real time using a laptop. The camera gives a live feed from the car (robot) point of view and extract images to be processed on the user’s computer. The camera has a Linux OS which runs a server with the objective of processing user’s instructions and sends the corresponding signals to the Arduino board. Once a complete instruction is received by the Arduino, it will execute the commands for motors control. The only sensor that the car has available is the camera. The program on the user’s computer runs in MATLAB, that calls the Java GUI and runs MonoSLAM [1] in parallel. The hardware setup is sketched in fig.1(b).

3 Autonomous Homing

We use MonoSLAM [1], so that the camera is the single sensor necessary to compute the pose of the vehicle. Basically, MonoSLAM consists of an Extended Kalman Filter (EKF) with Random Sample Consensus (RANSAC) embedded. The EKF makes a prediction of the vehicle’s pose using a constant velocity model. The update step is made by using the most voted hypothesis in RANSAC which generates hypothesis from candidate features matches. To detect local features in the images a corner extraction procedure is used. MonoSLAM is initialized assuming that an a priori probability distribution over the model parameters is known.

MonoSLAM provides an estimation of the pose of the car. One extracts from the vector \( state \) the position of the car

\[
W_C = state(1:3) = [x\ y\ z]^T
\]

and compute its rotation matrix in the world frame

\[
W_{RC} = q2r(state(4:7)).
\]

where \( q2r \) denotes the conversion for a quaternion to a rotation matrix. Defining \( c_2 = \sqrt{W_{RC}(1,1)^2 + W_{RC}(1,2)^2} \), i.e. the cosine of the rotation angle over the \( z \) axis, one has the the robot orientation angle is

\[
\theta_2 = \arctan2(-W_{RC}(1,3), c_2).
\]

MonoSLAM also gives us the inverse depth coordinates of every feature, from which we can compute the euclidean coordinates in the world reference frame.

The program saves the car’s position and orientation in every iteration in order to backtrack (home) to its initial position by the same way it reached its actual position. The backtracking is triggered by the user at any point he wants.

After the user activates the homing procedure, the car reverses and the controller actions take place. To calculate the vehicle’s pose in each iteration of homing, a simulator using the vehicle’s kinematic is being used where its reference orientation its calculated as

\[
\theta_{ref} = \arctan2 \left( \frac{Y_{ref} - Y_{actual}}{X_{ref} - X_{actual}} \right)
\]

This strategy is being used in real-time experiments, once the features matching in Mono SLAM running in MATLAB fails with abrupt movement.

4 Experiments and Results

In order to have an idea of the errors magnitude when the car moves, we have chosen the hard case where the car has just Forward Motion. As you can see in fig. 2(b) and fig. 2(c), the Mono SLAM acknowledges that the car is moving straight and the orientation is no different from the real one. When a considerable number of features moves its position on the image plan from one frame to the next, the algorithm understands the movement and realizes if the camera is sliding right, left, up, down, or if it is actually moving forward or backward. That can be verified on fig. 2(b) and fig. 2(c). As we can see in fig.2(d), the difference between ruler and monoslam position estimate is very large.

In a second experiment the car moves in a straight line and some few meters later, while still reading the pose from the MonoSLAM with abrupt movement.
request is issued. The difference between real and ours program pose estimation is shown in table 4. As you can see in table 4, the results are considerably different. The final vehicle’s pose is shown in fig.2(e).

From the experiments made, we can see that MonoSLAM easily detects changes in direction of motion, but it’s not very accurate at determining the magnitude of the camera’s forward motion, especially if there is no significant movement in more directions. The high computational cost of running MonoSLAM in Matlab makes the job at hand difficult. No abrupt movements can be done as otherwise the difference between two consecutive images will lead to no features matches. Using the kinematic model as the only prediction of pose while homing is possible, but leads also to a rapid increase of the pose prediction errors.

5 Future Work

As was already expected, visual odometry introduces error accumulation, even for short trajectories. One possible way to reset error accumulation is by modifying the scenario either by adding landmarks or fixed cameras informing the car location. An alternative way is to reset the error accumulation by comparing the actual image with the ones stored in an images-map [3], which is more interesting as the scenario is not required to be changed.

In an images-map based approach we are considering saving just image features that are registered (and 3D-reconstructed) along time. In each homing iteration we compute the camera matrix \( P \) from the 3D points stored previously and their current images. Given \( P \), we extract the position \( t \) and orientation \( R \) of the camera with respect to a global coordinate system [4]. First we apply QR factorization to the first three columns of \( P \), then transform from the QR to the RQ factorization and correct the sign of \( K \). In the end we obtain \( P = K[R | t] \). Figure 3 shows a simulated homing procedure encompassing calibration, pose computation and control.

6 Conclusion

This work started with the integration of the various hardware and software components into a mobile robot. This is still an ongoing project. Various positive conclusions can be extracted already. The first conclusion is that wireless IP cameras are practical and functional tools to help building teleoperated robots. Another conclusion is that it is possible and plausible to use odometry measurement using MonoSLAM but, as expected, we have considerable error accumulation. Further work is necessary to update our strategy, from the Mono SLAM robustness to a better homing procedure. As a general conclusion, we can say that is already possible to build a teleoperated vehicle with autonomous features at a reasonable cost.

Acknowledgments

This work has been partially supported by the FCT project PEst-OE / EEI / LA0009 / 2013, by the FCT project PTDC / EEACRO / 105413 / 2008 DCCAL, and by the FCT project EXPL / EEI-AUT / 1560 / 2013 ACDC.

References

Assessment of reliability of cerebrovascular reactivity measurements using breath-holding fMRI

Joana Pinto1
Inês Sousa1,2
Pedro Vilela3
Patricia Figueiredo1

1Institute for Systems and Robotics, Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Portugal
2Siemens SA, Healthcare Sector, Oporto, Portugal
3Imaging Department, Hospital da Luz, Lisboa, Portugal
email: joanasequeirapinto@gmail.com

Abstract

Functional magnetic resonance imaging (fMRI) is commonly used to measure cerebrovascular reactivity (CVR) as the response to breath-hold (BH) challenges. In this study, we aim to investigate the test-retest reliability of BH-fMRI CVR measurements using intra- and inter-subject coefficients of variation (CV) as well as intra-class correlation (ICC) metrics. Using a sinusoidal modeling approach, we found good reliability for both the amplitude and delay of CVR maps.

1 Introduction

The brain has autoregulation mechanisms that adjust regional cerebral blood flow (CBF), despite fluctuations in perfusion pressure, through changes in local vasculature caliber. In pathological conditions, this autoregulatory capacity can be compromised; therefore it is extremely useful to be able to evaluate it in order to obtain a marker for such diseases. This can be accomplished by assessing regional CVR. CVR is usually assessed by employing a vasoactive challenge to brain vasculature and measuring the associated CBF changes using an appropriate imaging modality. Blood Oxygen Level Dependent (BOLD) is a non-invasive MRI technique based on the fractional change in the local deoxyhemoglobin concentration and it is commonly used to measure CVR as the response to vasoactive respiratory challenges. The BH task is frequently used as the vasoactive stimulus. In this task subjects change their breathing pattern, by alternating periods of normal breathing with periods of breath-hold, leading to a hypercapnic condition (excessive levels of arterial CO₂ pressure, PaCO₂). Studies using BH hypercapnic tasks in conjunction with BOLD fMRI have successfully measure CVR in a non-invasive manner [1-3]. However, when studying a method it is important to evaluate the reliability of measurements and, on this matter, very few reliability studies of CVR and corresponding delay measurements have been conducted to date, with only one addressing both parameters using BOLD BH response [1]. The aim of this study is to calculate the parameters amplitude and delay of the BOLD response to a BH task and determine inter and intra-subject reliability of the voxelwise measurements using two reliability metrics: the coefficient of variation (CV) and the intra-class correlation coefficient (ICC).

2 Material and Methods

2.1 Data Acquisition

A group of 10 healthy subjects was studied on two different sessions on a 3T Siemens Verio MRI system. Two subjects were excluded due to excessive motion. BOLD images were obtained using GE-EPI (TR/TE=2500/50ms). The BH paradigm comprised three 75s cycles of 20s of BH preceded by inspiration and 5s of free breathing. Throughout the experiment end-tidal CO₂ pressure (PetCO2), a non-invasive surrogate of PaCO₂, was monitored using a capnograph.

2.2 Data Analysis

Pre-processing and statistical analysis

Image analysis was carried out with FSL (FMRIB’s Software Library, www.fmrib.ox.ac.uk/fsl) and MATLAB routines. Functional data were brain-extracted, motion-corrected and spatially smoothed (FWHM=5mm). Time series analysis was performed following a General Linear Model (GLM). BH BOLD response was modeled by a sinusoid with the paradigm frequency (1/75s), amplitude, Amp, and phase, Delay. This can be described by a linear combination of a sine and a cosine in the GLM [4].

\[
a_{12}\sin(\omega t) + a_{21}\cos(\omega t) = a_1\sin(\omega t + \phi)
\]

Eq. 1

The Amp parameter corresponds to the percent signal change relative to BOLD baseline, and can be calculated using:

\[a_1 = \sqrt{a_{11}^2 + a_{21}^2}\]

Eq. 2

The Delay is computed as:

\[\phi = \arctan\left(\frac{a_{21}}{a_{11}}\right) \quad \text{if } a_{11} \geq 0
\]

\[\phi = \arctan\left(\frac{a_{21}}{a_{11}}\right) + \pi \quad \text{if } a_{11} < 0\]

Eq. 3

Motion parameters were added to the GLM. Voxel-specific Amp and Delay maps were obtained from the cluster-thresholded (cluster p<0.05 and voxel Z>2.3) F test on the sine and cosine GLM coefficients. For each subject and session, the mean Amp and Delay were computed across all voxels exhibiting significant BOLD responses. Removal of outliers was done using specific thresholds: Amp < mean+3SD; Delay < 20s and mean-3SD < Delay < mean+3SD. The Amp values were normalized by the mean PetCO₂ change (nAmp). Delay and Amp maps were registered to MNI space.

Reliability metrics

CV of a measurement is defined as [1]:

\[CV = \frac{\text{std}(a)}{\text{mean}(a)} \times 100\]

Eq. 4

where a is the measurement. Reliability of measurements within subjects was computed as the intra-subject CV (CV_intra):

\[SD_w = \sqrt{\frac{\sum_{i=1}^{n}(a_i - \bar{a}_i)^2}{2 \times n}}\]

\[CV_{\text{intra}} = \frac{SD_w}{\mu} \times 100\]

where \(n\) is the number of subjects, \(a_i\) is the measurements of subject i on session 1 and 2, respectively and \(\mu\) is the mean value of measurements across sessions and subjects [5, 6]. Variability of measurements between subjects (CV_inter) was also assessed for each session using:

\[CV_{\text{inter}} = \frac{\text{std}(a_i)}{\text{mean}(a_i)} \times 100\]

Eq. 7

where \(a_i\) corresponds to the measurements of subject i. The final CV_inter was computed as the mean of the CV_inter of the two sessions[1]. CVs of Amp and Delay were computed for the whole brain (mean of all voxels measurements) and voxelwise.

A two-way random ICC between measurements was chosen to assess reliability of measurements using the ICC metric [7]:

\[ICC(2,1) = \frac{BMS - EMS}{BMS + (k-1)EMS + \frac{k(jMS-EMS)}{n}}\]

Eq. 8
where $BMS$ corresponds to the between-subjects mean square, $WMS$ to the within-subjects mean square, $EMS$ to the residual mean squares, $k$ the number of sessions and $n$ the number of subjects. A variation of ICC, ICCv, was used to measure the amount of total variance that can be explained by the intra-voxel variance. This type of ICC tests the consistency between sessions that measure ICC values indicate good intra-subject consistency of the spatial distribution of the BOLD signal for each individual [8, 9].

3 Results

The $Amp$ and $Delay$ maps obtained in this study, averaged across subjects and sessions, are shown in Fig.1.

![Group average Amp and Delay maps in MNI space. Only voxels with significant responses in at least 4 datasets are shown.](image)

Figure 1: Group average Amp and Delay maps in MNI space. Only voxels with significant responses in at least 4 datasets are shown.

Subject’s $Amp$ and $Delay$ values for each session are shown in Fig.2.

![Amp and Delay values averaged across brain for each subject in each session. Error bars represent standard deviations across brain.](image)

Figure 2: Amp and Delay values averaged across brain for each subject in each session. Error bars represent standard deviations across brain.

The results obtained with the different reliability metrics are presented in Table 1. The obtained CVs values are below the upper fiducial limit for acceptable variability in a normal distribution (33%) [1]. ICC values indicate good and excellent reliability in Amp and Delay measurements, respectively. Nevertheless, ICCv results demonstrate that, on average, the spatial distribution of Delay measurements is less consistent between sessions than that of the Amp measurements.

![CV inter, mapping (voxelwise) (Figure 3) demonstrates variations between brain areas across subjects that cannot be appreciated when using only whole brain metrics. CV inter maps were also calculated but no significant heterogeneities were observed across the brain (not shown).](image)

Table 1: Reliability metrics of CVR Amp and Delay measurements.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV inter</td>
<td>Amp</td>
<td>25.1%</td>
</tr>
<tr>
<td></td>
<td>Delay</td>
<td>21.1%</td>
</tr>
<tr>
<td>CV intra</td>
<td>Amp</td>
<td>16.3%</td>
</tr>
<tr>
<td></td>
<td>Delay</td>
<td>6.3%</td>
</tr>
<tr>
<td>ICC</td>
<td>Amp</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Delay</td>
<td>0.91</td>
</tr>
<tr>
<td>(mean across subjects)</td>
<td>Amp</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Delay</td>
<td>0.35</td>
</tr>
</tbody>
</table>

CV inter mapping (voxelwise) (Figure 3) demonstrates variations between brain areas across subjects that cannot be appreciated when using only whole brain metrics. CV inter maps were also calculated but no significant heterogeneities were observed across the brain (not shown).

![CV inter maps of Amp and Delay parameters.](image)

Figure 3: CV inter maps of Amp and Delay parameters.

4 Conclusions

The non-invasive BH fMRI protocol, combined with the proposed analysis strategy, provided both amplitude and delay measures of CVR on a voxel-by-voxel basis, with good inter- and intra-subject reliability using both metrics. Further investigation is needed to clarify inconsistencies in delay spatial patterns. Nevertheless, our results demonstrate that BH in conjunction with BOLD fMRI is a reproducible method to assess global CVR.

Acknowledgements

This work was supported by FCT projects PTDC/BBB-IMG/2137/2012 and PEst-OE/EEI/LA0009/2013.

References


A Critical Analysis about a Motion-based Approach to Extract Global Trajectories

Eduardo M. Pereira
epm@inesporto.pt
Jaime S. Cardoso
jaime.cardoso@inesporto.pt
Ricardo Morla
ricardo.morla@fe.up.pt

INESC TEC and Faculdade de Engenharia
Universidade do Porto, Porto, Portugal

Abstract
This paper summarizes a motion-based framework to extract global trajectories on unconstrained videos with cluttered scenarios [2]. That work was motivated by a challenging scenario of a retail shopping store. The aim was to extract global trajectories on a predetermined temporal gap in order to detect and recognize motion patterns. In detail, the two proposed motion tracking algorithms are compared and evaluated in our specific scenario. Future directions for improvements are pointed out, as well as some considerations for their implementation.

1 Introduction
Almost any public space has a CCTV (Closed Circuit Television) system installed due to the decreasing costs of video surveillance equipment and to the increasing demand of security in public areas. The excessive amount of collected information leads to the implementation of (semi-)automatic systems that can interpret crowded scenes in real domains in order to monitor individuals and their activities, detect common motion patterns, and identify unusual behaviors. This present various level of challenges, especially in unstructured motion scenes, where highly irregular motion patterns emerge from the motion of pedestrians with severe occlusions. Retail scenarios belong to this type of application context.

Understanding customers’ behavior and their purchase decision processes in an automatic way is an inestimable commercial advantage for the retail market. Such information can be inferred through the extraction of trajectories and the analysis of their motion pattern representations. Our scenario represents a physical structured scene with an unstructured appearance models changes abruptly from frame to frame due to the low of trajectories and the analysis of their motion pattern representations. The presented framework has a number of parameters that affect its track-

2 Motion Flow Framework
Under the aforementioned video conditions, we realize that motion in terms of flow vectors is the most informative cue for our dataset. As such, the framework combines the principle of instantaneous motion field with sampling strategy to extract meaningful flow vectors. The baseline framework, called Common steps (see Fig. 1), is composed by the sampling, motion flow estimation, and filtering steps. In particular: i) Sampling: this step extracts a set of interest points. The baseline considers a sparse sampling strategy, namely the FAST algorithm; ii) Motion Flow Estimation: this step is obtained from existing optical flow algorithms and is computed independently. The resulting motion flow map is used in the filtering step process in conjunction with the sampling points. The baseline considers the large-descriptor matching in variational model (LDOF) algorithm [1]; iii) Filtering: this step consists in filtering the flow vector of each sampling point at each frame. The baseline adopted a median filtering, with kernel size of $K = (5,5)$, that removes impulse noise, preserves edges, and smooths points in dense optical flow fields.

These steps form the common pre-processing block and are repeated for each frame. The output is a set of flow vectors at a specific frame, each one represented by $F_i = (x_i, y_i, u_i, v_i)$, where $(x_i, y_i)$ is the sampling point, and $(u_i, v_i)$ are the motion field components in $x$ and $y$ directions, respectively.

Before the motion tracking algorithm, a pre-processing step called Common-Block is necessary. It divides the image region by a regular spaced grid. Each region is denominated a cell, $C_i$, and contains the flow vectors positions that lay inside it (the baseline adopts a cell size of $C = (13,13)$). Each flow vector is encoded by $F_i = (x_i, y_i, M_i, \theta_i, t_i)$, where $(x_i, y_i)$ is the sampling point, $M_i$ is the flow magnitude, $\theta_i$ is the flow angle relative to positive $x$, and $t_i$ is the frame. A dual-threshold on flow magnitude is used to remove flow vectors that have extremely small and high motion information. On each cell $C_i$ a two-step hierarchical clustering approach is applied: i) Orientation quantization: considers a full-orientation histogram with 8 bins to express the orientation groups. The groups with weight above the histogram’s median value are taken into account for next clustering step; ii) Spatial clustering: consists on a spatial clustering on each orientation group. A k-means approach with center initialization was adopted. At the end, we get several clusters for each orientation group that are ordered in a descend-ent-weighted way, considering the number of flow vectors that belong to them. These groups represent the local dominant flows, which are described by $L_i = (x_i, y_i, n_i, \theta_i)$, where $(x_i, y_i)$ is the average position, $n_i$ is the total number of flow vectors, and $\theta_i$ is the average orientation angle. This block collects flow vectors every frame and at the end of a predefined temporal gap it executes the cluster resolution technique to fed, with the dominant flow groups’ cells, the subsequent motion flow tracking algorithm.

The presented framework has a number of parameters that affect its tracking results. The so-called Common block introduces a factor denominated cluster resolution, which is related to the two-step hierarchical clustering and reflects the decision for each cell’s orientation group to account with the most local dominant flow or all local dominant flows. The baseline considers the most local dominant flow as a good tradeoff between robustness and computational effort.

3 Critical Evaluation

Individually, each motion flow tracking algorithm introduces more parameters: 1) Grid-based algorithm increments two factors related to the scanning process: i) neighborhood similarity, reflects about the proximity between two local dominant flows and varies from the two pair-wise metrics used; ii) neighborhood continuity, reflects about the integration of a different orientation group at the final iteration in the continuity decision; 2) Kernel-based algorithm presents three more factors related to the sink-seeking process: i) bandwidth initialization, is related to the ini-

C
An evaluation methodology was designed to cluster trajectories by similarity to obtain the common ones and to measure correspondence between extracted and annotated trajectories. For clustering, we use several distance functions, that account with trajectory position and shape, to build a similarity matrix to be used as input in a spectral clustering that has a K-means technique on its final step. For more details, refer to [2]. We report the quality of correspondence between the most similar extracted trajectory and annotated trajectory with the miss detection and false positive rates. This procedure uses an one-against-all distance function between each annotated trajectory and all auto-generated ones, to obtain a distance matrix and solves the assignment problem with the Hungarian algorithm. The miss detection rate is the number of unmatched annotated trajectories, and the false-positive rate is the number of annotated trajectories matched minus the correctly classified matches.

Statistical values were computed from annotated trajectories and automatic ones, obtained from both algorithms, to describe their shape information, evaluate scene complexity, and compare results. Results are reported on the following tables.

Table 1: Characteristics of trajectories jectories.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Auto</th>
<th>GT-Grid</th>
<th>GT-Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>786.18</td>
<td>46.36</td>
<td>13.04</td>
</tr>
<tr>
<td>Width</td>
<td>20.34</td>
<td>4.28</td>
<td>1.55</td>
</tr>
<tr>
<td>Height</td>
<td>482.32</td>
<td>29.95</td>
<td>1.84</td>
</tr>
<tr>
<td>Inclination</td>
<td>10.38</td>
<td>74.48</td>
<td>67.68</td>
</tr>
<tr>
<td>Incidence</td>
<td>70/109</td>
<td>39.56</td>
<td>37.68</td>
</tr>
<tr>
<td>Max</td>
<td>106.12</td>
<td>19.02</td>
<td>10.9</td>
</tr>
<tr>
<td>Min</td>
<td>19.32</td>
<td>2.28</td>
<td>1.55</td>
</tr>
<tr>
<td>Mean</td>
<td>44.27</td>
<td>39.56</td>
<td>19.02</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>927.52</td>
<td>48.86</td>
<td>48.86</td>
</tr>
<tr>
<td>Correlation Coefficient</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
</tr>
</tbody>
</table>

We verify that manual trajectory set is more complex and less sparse than any of the auto-generated trajectory set. In general, trajectories extracted from the Kernel-based method are smoother and shorter than the ones extracted from the Grid-based method, which is more sensitive to noise and does not keep track of a smooth flow. This is explained because the kernel window does not accept flow vectors with large opposite directions, therefore the sink-seeking process maintains a coherent flow and its sliding window’s overlap permits to build smoother trajectories with greater points. The Grid-based method produces more trajectories than the Kernel-based method, 339 trajectories against 192. It is related to the short continuity of the scanning process involving both neighborhood continuity and neighborhood similarity effects, whose also affect flow coherence.

To infer trajectory parameters similarity between auto-generated trajectories and annotated trajectories we use histogram comparison. The results are reported on Table 4, where we verify that the Kernel-based method is more correlated on length (ℓ) and number of points (n_p) parameters, Grid-based method has a strong shape divergence (ζ_{divergence}) correlation, and mostly completely correlation on shape complexity (ζ_{complexity}) parameter. For the remainders, both methods have equal degree of correlation.

Through qualitative inspection of clustering for common trajectories (refer to [2]), we verify good partitioning results using several distance metrics, whose lead us to conclude that the number of common trajectories extracted from both algorithms is similar to the ground truth data, even with different number of trajectories. We use the same trajectory distance measures for both clustering and assignment steps, and we report the most significant results of the assignment procedure on Table 5.

Table 2: Characteristics of annotated trajectories from motion tracking algorithms.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Auto</th>
<th>GT-Grid</th>
<th>GT-Kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>786.18</td>
<td>46.36</td>
<td>13.04</td>
</tr>
<tr>
<td>Width</td>
<td>20.34</td>
<td>4.28</td>
<td>1.55</td>
</tr>
<tr>
<td>Height</td>
<td>482.32</td>
<td>29.95</td>
<td>1.84</td>
</tr>
<tr>
<td>Inclination</td>
<td>10.38</td>
<td>74.48</td>
<td>67.68</td>
</tr>
<tr>
<td>Incidence</td>
<td>70/109</td>
<td>39.56</td>
<td>37.68</td>
</tr>
<tr>
<td>Max</td>
<td>106.12</td>
<td>19.02</td>
<td>10.9</td>
</tr>
<tr>
<td>Min</td>
<td>19.32</td>
<td>2.28</td>
<td>1.55</td>
</tr>
<tr>
<td>Mean</td>
<td>44.27</td>
<td>39.56</td>
<td>19.02</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>927.52</td>
<td>48.86</td>
<td>48.86</td>
</tr>
<tr>
<td>Correlation Coefficient</td>
<td>0.9996</td>
<td>0.9996</td>
<td>0.9996</td>
</tr>
</tbody>
</table>

4 Conclusions and Future Approach

Our framework incorporates two motion flow tracking algorithms that present good and promising results under special video conditions, where conventional crowd motion and multiple tracking approaches are not useful. The future step involves the validation of exact metrics to evaluate the influence of algorithm’s parameters, the integration of an automatic process to select the best algorithm’s parameters, and the inclusion of temporal information. A fluid-dynamics technique could be an interesting approach to integrate motion through temporal domain in order to extract several and meaningful global flow representations that encode spatial and temporal changes in the scene, as well as local statistics on the vectors around each trajectory point in the form of discrete distributions. We believe that such representations reveal new forms to recognize, identify and learn motion behavior, and detect related events on specific context scenarios.

Acknowledgements

This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281, and for the PhD grant with reference SFRH/BD/51430/2011.

References


Ground-plane based indoor mobile robot localization using RGB-D sensor

Miguel Vaz
miguelvaz@isr.ist.utl.pt
Rodrigo Ventura
rodrigo.ventura@isr.ist.utl.pt

Abstract

This paper addresses the problem of absolute localization in an indoor environment using a RGB-Depth camera. The approach is based on the use of the ground region perceived by the RGB camera to detect and decode its position and edges. The localization system uses this data to match it with a known on-board map. The ground plane detection algorithm is designed to be robust to vibration or disturbances during the robot motion. The localization system is based on the particle filter, fusing odometry with ground region matching where each particle’s weight is proportional to the quality of correspondence between the ground edge estimation and the nearest walls. Promising results were obtained and are presented in this article.

1 Introduction

In the past years, several publications have addressed the localization problem based on wall detection methods [1]. These systems have some limitation, especially in unstructured environments where walls are sometimes difficult to detect (hidden by furniture or not present). Another issue is that this algorithm can be mislead by a planar obstacle. We tackled the localization problem using the RGB-Depth camera and the Particle filter method. The data acquired by the RGB-Depth camera form a point cloud, which consists in a set of 3D points \((x, y, z)\) in the camera frame. Each 3D point of the set is associated to one pixel on the image plane of the camera. In this work we used the Microsoft Kinect sensor. Since the localization system is based on the ground observation, we assumed that the camera is always pointing down with the floor on a big part of the FOV. First we detected the ground point cloud based on a pre-calibrated ground plane and on a dynamic threshold filter. The filter point cloud edges are then estimated and used to calculate the particle weight on the Particle filter algorithm [5]. The weight of each particle is inversely proportional to the distance of the edges from the nearest wall, as seen by the particle position. While the localization system developed by Biswas and Veloso [1] is based on wall detection and random sampling of the depth images for plane detection, our approach uses ground detection algorithms that are less sensible to planar obstacles and usable in unstructured environments.

2 Floor Detection

2.1 Floor detection

The floor is modeled as a plane and parametrized according to the following normalized equation [4]:

\[ ax + by + cz + d = 0, \]  
(1)

where \(a, b, c\) is the normal vector, \(d\) the distance to the origin and \((x, y, z)\) are the coordinates of a point on the ground in the camera frame \((X, Y, Z)\). Before starting the localization system we performed a calibration of the floor. This calibration setup is shown in Fig. 1 where a chess board pattern is placed inside the camera’s FOV, maintaining the robot still. A corner detection algorithm [6] is used on the colored image to estimate the pixel coordinates of the board’s inner corners. With the obtained data and the corresponding depths provided by the camera, we estimated the 3D coordinates of the chess corners.

The calibrated floor parameters \(a', b', c', d'\) are estimated by applying the Least Squares on the estimated 3D inner corners position, providing an approximated ground model. A new frame defined by the orthonormal basis \((X_p, Y_p, Z_p)\) is associated to the calibrated ground plane as shown in Figs. 1 and 3, where \(X_p\) axis is the normalization of the \(Z\) axis’ projection and the \(Z_p\) axis is the normal vector of the plane. The \(Y_p\) axis results from the external product of \(X_p\) with \(Z_p\). This definition assumes that the \(Z_e\) axis points forward the robot.

To estimate the ground model during the robot’s movements, a dynamic threshold function is defined to remove from the complete point cloud the points that are further away from the calibrated ground plane. The threshold function is defined as two planes, one down and the other one up symmetric along a pre-calibrated ground plane model. The function value grows as we walk away from the robot, as one can see in Fig. 2 where in black we have the pre-determined ground plane position, in orange the true position of the plane and in dashed yellow the value of the threshold function. Therefore a point which distance from the calibrated plane is bigger then its corresponding threshold value is discarded (like point \(P_1\) in Fig. 2).

![Figure 1: Calibration setup with the FOV (green) and ground plane intersection with FOV planes (red).](image1)

![Figure 2: Threshold dynamic filter. Its values depends on the point distance to the floor frame.](image2)

The resulting filtered point cloud is formed by the points of the ground and by some outliers, in particular from the obstacles that are further from the robot. To remove these outliers a random sample consensus (RANSAC) algorithm [3] is applied. The result is a point cloud formed just by the floor point cloud and the real ground plane parameters. An example of such point cloud is illustrated in Fig. 3.

We developed this floor detection algorithm because the calibration process by itself is not enough to detect the floor point cloud on-line in a robust way because of the vibration during the robot’s motion. This makes the floor parameters change significantly comparing with the calibrated parameters.

2.2 Edges estimation

After the floor detection, an edges estimator was designed to detect the edges of the ground seen by the robot. For that we applied the concave hull algorithm [2]. The result is a list of points forming the polygon of the ground seen in the camera FOV that includes the edges created by the end of the FOV. A filter is then applied to remove them. As one can see in Fig. 1, the intersection of the FOV planes (green lines) with the ground plane is computed.
lines). Then we filtered the points of the polygon that are near the intersection of the two planes.

### 2.3 Cost function

A cost function is defined as the $L^1$ norm of the distances between the ground edges and the nearest walls according with a determined particle, as shown in the following equation:

$$ C(L, (x_j, y_j, \theta_j)) = \sum_{i=0}^{n} D(P_{L}^{(x_j, y_j, \theta_j)}) $$

where $L$ is the list of edge points, $(x_j, y_j, \theta_j)$ are the coordinates of the particle $j$, $D(\cdot)$ is the distance from a point to the nearest occupied pixel and $P_{L}^{(x_j, y_j, \theta_j)}$ is the $P^{th}$ point in the edges list $L$ transformed to the $(x_j, y_j, \theta_j)$ coordinates, i.e. the edges as seen in the $j$ particle position.

As an illustration, for the list $L$ obtained looking at a wall’s corner and fixing the robot orientation, one can see the function value for a wide area near the true position of the robot.

![Figure 4: Cost function defined, fixing the orientation, in the red area in Fig. 5 for the blues edges detected (wall corner).](image)

If one tries to estimate the robot position by the minimum of the function, we find that it is consistent with the robot true position as one can see in the Fig. 5 by the good correspondence of the walls (black points) with the edges detected (blue points) at the position estimated by the minimum (magenta cross).

### 3 Robot Localization System

The localization system is formed by the particle filter [5] where the particles’ weight is inversely proportional to the particle cost. The particle cost is estimated according to the previous defined cost function. This way, the system at the predict step moves the particles according to the odometry readings and adds zero mean Gaussian noise. At the update step, the system gets the point cloud from Kinect, estimates the floor point cloud, performs the edges detection and calculates each particle’s cost value. Then the resulting particle set of the resampling based on the particle weight, i.e inverse of the cost, is used on the next iteration of the particle filter.

![Figure 5: Area where the cost function was calculated (red lines), the position of less cost value (magenta cross) and the measurements in the global frame as seen in the minimum cost position (blue lines).](image)

### 4 Results

As a preliminary result, we ran the particle filter with the robot in a fixed position with a start particle set of 300 uniformly distributed around the real pose of the robot in a 4 by 4 square meter and with $\pm 10^\circ$ of orientation range. The system successfully managed to convert to the real position as one can see in Fig. 6.

![Figure 6: Result of the particle filter algorithm with the ground detection (red points) and edges estimation (blue lines).](image)

### 5 Conclusions and Future Work

In this paper we described a localization system based on a RGB-Depth camera on an indoor environment. The proposed implemented system appears to converge with small errors for the scenario specification used in the experiment. As a future work, we would like to perform further tests on the system in different scenarios and carry out a deeper analysis of the errors in the system including during robot motion.

### References


Abstract
The simulation of quadrotor dynamics is an essential tool for the development of autonomous behaviours for these platforms, since they prone to physical damage upon crashing. However, validation of these behaviours require the simulation to be faithful. Dynamical models for quadrotor can be formulated as a state space, continuous time, differential equations, depending on a set of physical parameters of the platform. These parameters are the mass, inertia tensor, and propeller thrust factor. This paper addresses the problem of estimating these parameters from flight data. A Least Squares (LS) approach is taken for the estimation. The estimation problem is formulated here, together with preliminary results. The approach is validated with simulated data. We also discuss how the presence of noise hinder the correct estimation of parameters.

1 Introduction
Developing simulators is very common in the field of robotics, regarding the quad rotors is a crucial step due to the risk of physical damaging in real flights. To have a reliable simulator is essential a good estimation of the parameters. In this paper, is presented a quadrotor dynamical model as well as a method to estimate his parameters, using specific flight data in order to simplify the LS estimations.

The LS is a widely used method and converges for the true value of the parameter [1], but as shown in [3], this is only true when the input data is noise free. In systems where the input and output are corrupted with noise, the estimation will be bias. Hence the method is tested using a simulator, in a first approach only the output data is corrupted. Secondly both output and input are corrupted with noise in order to verify the influence in the estimations.

2 Quad-rotor dynamic model

![Quadrotor body-fixed frame](image)

The quad-rotor model is fully studied in [2]. In this paper it is just presented the following set of equations

\[
\begin{align*}
\dot{p} & = \frac{I_{yy} - I_{zz}}{I_{xx}} qr + \frac{U_2}{I_{xx}} \\
\dot{q} & = -\frac{I_{xx} - I_{yy}}{I_{zz}} pr + \frac{U_3}{I_{zz}} \\
\dot{r} & = \frac{I_{xx} - I_{yy}}{I_{zz}} pq + \frac{U_4}{I_{zz}}
\end{align*}
\]

(1a)

(1b)

(1c)

\[
\begin{align*}
U_1 &= b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) \\
U_2 &= bI(\Omega_2^2 - \Omega_1^2) \\
U_3 &= bI(\Omega_3^2 - \Omega_1^2) \\
U_4 &= d(\Omega_2^2 + \Omega_3^2 - \Omega_1^2 - \Omega_4^2)
\end{align*}
\]

(2a)

(2b)

(2c)

(2d)

Where \( \dot{\omega} = [p, q, r]^T \) is the angular accelerations and \( \omega = [p, q, r]^T \) is the angular velocities with respect to \( x, y, z \) in the body-fixed frame. \( \Omega = [\Omega_1, \Omega_2, \Omega_3, \Omega_4] \) is the propellers’ speed and Throttle\( (U_1) \), Roll\( (U_2) \), Pitch\( (U_3) \) and Yaw\( (U_4) \) are the four possible movements of the quad-rotor.

The parameters that need to be estimated are the thrust \( (b) \), drag \( (d) \) and the inertia tensors \( I_{xx}, I_{yy} \) and \( I_{zz} \).

The distance between the center of the quad-rotor and the center of the propellers, \( l \), is a known parameter.

3 Parameters Estimation
In this section a method is proposed which aims to estimate the parameters individually, using flight data with specific movements in order to simplify the equations presented in section 2.

3.1 Thrust estimation
To estimate \( b \), only equation (2a) is used. Assuming that the quad-rotor is hovering, \( U_1 = mg \), where \( m \) is the quad-rotor mass and \( g \) the acceleration due to gravity. Using this assumption

\[
\frac{Y}{mg} = \frac{X}{b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2)}
\]

(3)

and \( b \) LS estimation [1] is

\[
\hat{b} = (X'X)^{-1}(X'Y)
\]

(4)

In actual flights it could be tricky to keep the quad-rotor hovering, therefore another solution is proposed. Using Newton’s second law over \( z \) axe

\[
F = m\ddot{z} = b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) - mg
\]

(5)

Doing a double integrating on equation (5) and considering that the final and initial positions are the same, \( b \) is deducted as follows

\[
\int \int_{t_1}^{t_2} (m\ddot{z}) = \int \int_{t_1}^{t_2} (b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) - mg)
\]

\[
m(\ddot{z}(t_2) - \ddot{z}(t_1)) = b \int \int_{t_1}^{t_2} (b(\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2) - mg)
\]

\[
\hat{b} = \frac{mg \int \int_{t_1}^{t_2} (1)}{\int \int_{t_1}^{t_2} (\Omega_1^2 + \Omega_2^2 + \Omega_3^2 + \Omega_4^2)}
\]

(6)

3.2 \( I_{xx} \) Estimation
To estimate \( I_{xx} \), flight data will only contain roll movements, which implies that \( \omega = [p, 0, 0]^T \). Applying this, equation (1a) becomes

\[
\dot{p} = \frac{bI(\Omega_2^2 - \Omega_1^2)}{I_{xx}}
\]

(7)

In order to perform the LS regression it is necessary to turn the equation into a discrete domain. To approximate the derivative, a forward finite difference is used.
\[
p(t + h) - p(t) = \frac{b(\Omega_2^2 - \Omega_1^2)}{I_{xx}} \quad (8)
\]

using \(b\) value estimated previously, \(I_{xx}\) LS estimation is given by

\[
\hat{X} = \frac{p(t + h) - p(t)}{bbh} I_{xx}
\]

\[
I_{xx} = (X'X)^{-1}(X'Y) \quad (9)
\]

3.3 \(I_{yy}\) Estimation

This case is similar to the previous one, but instead of roll movements, pitch movements are used, hence, \(\omega = [0, q, 0]^T\). Applying as before, to equation (1b) and doing a forward finite difference to approximate the derivative

\[
g(t + h) - g(t) = \frac{b(\Omega_2^2 - \Omega_1^2)}{I_{yy}} \quad (10)
\]

the \(I_{yy}\) LS estimation is

\[
\hat{X} = \frac{g(t + h) - g(t)}{bbh} I_{yy}
\]

\[
I_{yy} = (X'X)^{-1}(X'Y) \quad (11)
\]

3.4 \(I_{zz}\) Estimation

To estimate \(I_{zz}\) the equations (1a) and (1b) are used, in this case the flight data will contain the three movements Roll, Pitch and Yaw. Using the forward finite difference and the previous estimations

\[
\frac{p(t + h) - p(t)}{h} I_{xx} + qr I_{yy} + b(\Omega_1^2 - \Omega_1^2) = \frac{X}{\Omega_1} \quad (12)
\]

\[
\frac{q(t + h) - q(t)}{h} I_{yy} + pr I_{zz} - b(\Omega_2^2 - \Omega_2^2) = \frac{Y}{\Omega_2} \quad (13)
\]

the \(I_{zz}\) LS estimation is

\[
I_{zz} = \begin{bmatrix} x_1^\top & x_1^\top \\ x_2^\top & x_2^\top \end{bmatrix}^{-1} \begin{bmatrix} x_1^\top \\ x_2^\top \end{bmatrix} \quad (14)
\]

3.5 Drag Estimation

The last parameter to estimate is \(d\), where the data will only contain Yaw movements and the equation (1c) is used. This results in \(\omega = [0, 0, r]^T\) and approximating the derivative by the forward finite difference and using the \(I_{zz}\) estimation

\[
\frac{r(t + h) - r(t)}{h} I_{zz} = \frac{X}{\Omega_2} \quad (15)
\]

the \(d\) LS estimation is

\[
\hat{d} = \frac{1}{(X'X)^{-1}(X'Y)} \quad (16)
\]

4 Results

To test the method, a simulink simulator developed in [2] is used. A block is included to add white Gaussian noise to the variables, in order to exercise the sensitivity of the method. The simulator assumes the following

<table>
<thead>
<tr>
<th>Flight</th>
<th>LS</th>
<th>Double Int.</th>
<th>LS</th>
<th>Double Int.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.415e−5</td>
<td>5.427e−5</td>
<td>5.4195e−5</td>
<td>5.437e−5</td>
</tr>
<tr>
<td>2</td>
<td>5.4148e−5</td>
<td>5.427e−5</td>
<td>5.4180e−5</td>
<td>5.437e−5</td>
</tr>
<tr>
<td>3</td>
<td>5.418e−5</td>
<td>5.4268e−5</td>
<td>5.4186e−5</td>
<td>5.4379e−5</td>
</tr>
</tbody>
</table>

Table 1: Estimation values of parameter \(b\) with and without noise in \(X\)

<table>
<thead>
<tr>
<th>Flight</th>
<th>(I_{xx})</th>
<th>(I_{yy})</th>
<th>(I_{zz})</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.1e−3</td>
<td>8.0e−3</td>
<td>15.1e−3</td>
<td>1.0985e−6</td>
</tr>
<tr>
<td>2</td>
<td>7.9e−3</td>
<td>8.0e−3</td>
<td>13.9e−3</td>
<td>1.0968e−5</td>
</tr>
<tr>
<td>3</td>
<td>8.1e−3</td>
<td>8.1e−3</td>
<td>13.6e−3</td>
<td>1.101e−6</td>
</tr>
</tbody>
</table>

Table 2: Estimation values of parameters \(I_{xx}, I_{yy}, I_{zz}\) and \(d\) with and without noise in \(X\)

The measurements were taken with flights of 100 seconds and with a sampling period of 0.05 seconds. For each parameter were done six flights, where three of them were only corrupted with noise in the output data and the other three had noise in both output and input data.

Table (1) shows the estimations values of \(b\), both method presented in section 3.1 were tested. The two methods proved to be effective, even in the presence of noise in the input data.

Table (2) presents the estimations values for the remaining parameters, it can be observed that if the input data is error free the estimations are close to the true value, but in the scenario where both output and input are noisy, the estimation of \(I_{xx}\) and \(I_{yy}\) are bias [3] and the error propagates to the remaining estimations.

5 Conclusions

This paper presented a method to estimate the parameters of a quad rotor dynamic model, the preliminary results shows lack of precision of the method when both input and output data are corrupted with noise.

It is noteworthy that all inertia tensors can be estimated at the same time using equations (12) and (13) from section (3.4), but in that scenario the movements Roll, Pitch and Yaw must be executed simultaneously, which is a tricky flight to perform. Therefore it was chosen to estimate them separately for a better estimation of \(I_{xx}\) and \(I_{yy}\).

Future work consist in estimating the parameters with real data taking into account that a noise free input is needed, which can be difficult to achieve. Another approach is to make a non-parametric estimation of the system.

References

### Abstract

The ever-growing biological examples of functionally active non-coding elements, reveals the urgency in the search of such elements in disease contexts. The hereditary diffuse gastric cancer syndrome has been for long associated with genetic alterations in a particular gene, CDH1. However, many cases of HDGC remain without an adequate genetic diagnosis even though most of them display loss of expression of the corresponding protein, E-cadherin. The lack of a proper genetic diagnosis may be due to the fact that never was the non-coding portion of CDH1 (over 70% of the whole gene) screened in HDGC patients. Therefore, and using a combined bioinformatics and biological approach, we have designed an innovative strategy to uncover novel non-coding based variants for CDH1. We have used germline DNA of 90 HDGC probands without any of the classical mutations described for CDH1 to test our combined strategy. More than 500 novel variants were detected for the 90 HDGC probands, 110 of which rare (simultaneously present in 3 or less probands). From the 110, 47 were already subjected to PCR validation and 42 were proven to be true positives, showing that our strategy has a true positive rate of approximately 90%. We further designed a bioinformatics strategy to classify such true positive variants into increased likelihood of functional relevance using several criteria and those ranking higher will be promptly tested using several in vitro assays. The strategy herein described may unravel novel HDGC-related CDH1 variants, allowing for a more complete genetic screening of probands. Moreover, our strategy can be applied to other disease-causing genes, for which limited genetic screening and few information on non-coding variants is currently available.

### Introduction

Hereditary diffuse gastric cancer (HDGC) syndrome, although rare, is severe, highly penetrant, difficult to diagnose and incurable [1]. Germline E-cadherin (encoded by CDH1) coding mutations and large deletions explain 40% of the cases, but still HDGC tumours form ~90% of HDGC families display loss of protein expression [2,3]. In addition, large genetic screens have failed to identify germline defects in other relevant HDGC genes. Therefore CDH1-E-cadherin continues to be the main causal gene associated with HDGC syndrome.

Importantly, the classical CDH1 gene screening performed only assessed genetic alterations in the coding portion of the gene. In fact, DNA variations outside protein-coding sequences may underlie the next layer of gene expression regulation, with impact in health and disease. Thousands of well-annotated and conserved elements are known to exist in the non-coding genome and several have been already implicated in disease.

We hypothesized that a potential mechanism underlying the observed loss of E-cadherin expression in >90% of HDGC samples, is the presence of novel mutations occurring at CDH1 non-coding sequences, until now unscreened. Supported by these observations, Steenstorm et al. showed that complete deletion of CDH1 intron 2 abrogates initiation and maintenance of CDH1 expression, in several organs of the mouse embryo including stomach [4,5]. This work highlighted the role of regulatory elements, within CDH1 non-coding regions, as crucial factors for proper CDH1 expression.

In light of all these observations, the aim of our work was to identify novel CDH1 HDGC-related variants within the non-coding portion of this gene, using a combined biological and bioinformatics strategy, developed by our lab. Importantly, the strategy developed can also be applied to the non-coding genetic screening of other disease-related genes, enriching current genetic diagnosis.

### Results

To address our aim, we have screened 90 bonafide CDH1-negative HDGC probands by next generation sequencing of the complete CDH1 locus aiming at disclosing novel non-coding variants that may underlie HDGC syndrome.

The first step was to fractionate the entire human CDH1 locus (~70kb) in a given number of fragments taking into account the position of repetitive elements. Such elements are rarely screened given the high level of exact or nearly exact copies found throughout the human genome. Using Ensembl and RepBase databases [6,7] and private PERL scripts, we have fractionated the entire CDH1 locus into 80 overlapping fragments. Fragmentation was performed preventing the presence of any repetitive element within 50 bp of each fragment border. This allows: 1) the design of fragment-specific amplification primers and; 2) repetitive elements sequencing in the following steps (Figure 1).

Next, fragment-specific primers were designed, validated using control samples and optimized into a set of multiplex PCR reactions for a less time-consuming amplification. Germline DNA from the blood of 90 CDH1-negative HDGC probands was then used for fragment amplification and the corresponding amplicons (after a random-sample validation step) were amplified and sent to the BCCA Genome Sequencing Service in Vancouver, Canada. At this service, the complete set of amplicons from each proband was barcoded and sequenced, using an Illumina Genome Analyzer was performed (Figure 1). The resulting reads were aligned to the Human Genome (hg18) using the open-source software BWA [8] and SAM files produced later on converted in BAM files using SAMtools [9]. The BAM files, containing all uniquely mapped reads and corresponding genomic coordinates were next used for variant calling using the GATK software [10]. This analysis was performed by the BCCA Genome Sequencing Service and is part of their standard analysis pipeline. The resulting VCF files encompassed all non-reference-sequence variants detected in the set of amplicons from each proband sequenced. These variants were further selected using stringent quality criteria: only variants with a quality level above 3000 (directly related with single nucleotide coverage and is phred-scaled) and with at least 25% of the reads mapped to the variant were selected (Figure 1).

The resulting high-quality variants were then crossed with several online variant databases, such as dbSNP to separate annotated from novel variants [11]. In all, over 1000 variants were detected in the 90 proband cohort sequenced: 518 annotated variants and 647 novel variants. The presence of a large set of known variants strengthened our analysis strategy and provides valid insight into the population characteristics of the probands sequenced (ongoing analysis). The overall distribution of all detected variants is displayed in Figure 2.

Focusing on novel variants (n=647), only those considered rare were selected for PCR validation in the original proband blood DNA. The concept of rare was defined based on the characteristics of the HDGC syndrome: never was the same genetic alteration detected in more than 3 distinct families. Therefore, only novel variants present in 3 or less probands were further selected for PCR validation, resulting in 110 variants (Figure 1 and Figure 2). From the resulting 110 novel and rare CDH1 non-coding variants, 47 were already submitted for PCR validation. We have confirmed 42 out of the 47 to be true positives, rather than technical artefacts, constituting a rate of true positives of 90%.

Next, we ran a systematic bioinformatics approach to assess the potential impact of previously identified non-coding variants in CDH1 transcription control, using the software PROMO [12] and by comparing the reference sequence with the variant-encompassing sequence, we recurrently predicted the creation of novel transcription factor binding sites (TFBS) for well-known CDH1 expression repressors and the loss of TFBS for CDH1-positive modulators.

---

**Figure 1:** Combined bioinformatics (orange) and biological (blue) strategy used for detection of CDH1 non-coding variants in HDGC probands.

Next, fragment specific-primers were designed, validated using control samples and optimized into a set of multiplex PCR reactions for a less time-consuming amplification. Germline DNA from the blood of 90 CDH1-negative HDGC probands was then used for fragment amplification and the corresponding amplicons (after a random-sample validation step) were amplified and sent to the BCCA Genome Sequencing Service in Vancouver, Canada. At this service, the complete set of amplicons from each proband was barcoded and sequenced, using an Illumina Genome Analyzer was performed (Figure 1). The resulting reads were aligned to the Human Genome (hg18) using the open-source software BWA [8] and SAM files produced later on converted in BAM files using SAMtools [9]. The BAM files, containing all uniquely mapped reads and corresponding genomic coordinates were next used for variant calling using the GATK software [10]. This analysis was performed by the BCCA Genome Sequencing Service and is part of their standard analysis pipeline. The resulting VCF files encompassed all non-reference-sequence variants detected in the set of amplicons from each proband sequenced. These variants were further selected using stringent quality criteria: only variants with a quality level above 3000 (directly related with single nucleotide coverage and is phred-scaled) and with at least 25% of the reads mapped to the variant were selected (Figure 1).

<table>
<thead>
<tr>
<th><strong>Step</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Combined bioinformatics strategy includes: a) targeted DNA sequencing (Illumina Genome Analyzer), b) variant calling (VCF files with sequence and quality information).</td>
</tr>
<tr>
<td>2</td>
<td>Primer design and amplification specificity.</td>
</tr>
<tr>
<td>3</td>
<td>Multiplex reactions setup.</td>
</tr>
<tr>
<td>4</td>
<td>PCR validation.</td>
</tr>
<tr>
<td>5</td>
<td>Novel rare &amp; high-quality variants.</td>
</tr>
<tr>
<td>6</td>
<td>Annotation variants (n=647).</td>
</tr>
<tr>
<td>7</td>
<td>Novel variants (n=647).</td>
</tr>
<tr>
<td>8</td>
<td>Variant calling using bioinformatics tools.</td>
</tr>
<tr>
<td>9</td>
<td>Alignment of all reads to Human Genome (hg18).</td>
</tr>
<tr>
<td>10</td>
<td>Novel rare &amp; high-quality variants (n=110).</td>
</tr>
<tr>
<td>11</td>
<td>Alignment of all reads to Human Genome (hg18).</td>
</tr>
</tbody>
</table>

---

**Expression Regulation in Cancer, IPATIMUP, Porto, Portugal**

BCCA, Vancouver, Canada
Both medium (microarrays) and large scale genomic/transcriptomic approaches will be firstly tested and further confirmed, according to our strategy, an increased likelihood of functional relevance leading to the alteration of a TFBS for a classical CDH1 cadherin TF present, i.e., tumour samples datasets, genetically modified cell-lines, microorganisms; 
- Generate data supporting a hypothesis-driven research proposal: i.e., preparation of grant proposals, discussion section in papers and thesis, etc.

For more information visit Bioinf\*:Bio website: http://www.frombioinformatics2biology.com

**Conclusion**

The strategy described has led to the discovery of over 500 unannotated variants, with at least 42 putatively associated with HDGC syndrome. In the near future, the knowledge acquired with this ongoing project may enrich the classical genetic screening offered to HDGC probands and families, shedding a necessary light into this syndrome. Moreover, the strategy herein described may be applied to other disease-causing genes possibly broadening the genetic screening of the respective syndromes.

**References**

Abstract

Carotid plaque characterization is an open problem with very relevant clinical implications, because carotid atherosclerotic disease is one of the main causes for stroke and other life threatening conditions due to embolization and/or flow reduction.

In this paper we explore the usefulness of monogenic decomposition based features in the characterization of carotid plaques, and its usage in a classification approach to identify their symptomatic vs. asymptomatic condition.

Results show the high discrimination power the proposed features and their high classification performance in classifying symptomatic vs. asymptomatic plaques from ultrasound images and the potential for quantifying it’s vulnerability.

1 Introduction

The arteries that supply our brains, the carotids, are prone to develop atherosclerotic plaques that reduce blood flow. More dangerous than that, they are vulnerable to rupture or break-away and block smaller vessels causing ischemia (death) to the surrounding tissues. Carotid bifurcation disease is actually responsible for one-third of acute cerebrovascular events, hence it has a major clinical and social impact.

A stable carotid plaque is usually benign with a stroke risk around 3% annually, but a more vulnerable plaque might cause myocardial infarction, stroke and lower limb ischemia. Correct characterization of the carotid disease is vital for an accurate decision to surgically remove the plaque (carotid endarterectomy) or not. The major premise here is that a vulnerable plaque contains predictive information for future cardiovascular events. Hence, its detection might play a major role in the treatment decision that has important clinical, social and economical consequences.

Ultrasound (US) is a suitable imaging technique to assess this pathological condition mostly because it provides real-time visualization and interpretation of the carotid plaques, its non-invasive, does not involve ionizing radiation, it is cheap and is very common in clinical facilities.

When diagnosing plaque vulnerability, the degree of stenosis (arterial lumen narrowing) is up to now considered the most important feature. This metric, together with other patient information such as age, health and clinical history are usually used by clinicians to subjectively decide upon endarterectomy (surgery). But several research studies ([1, 2, 4, 6, 7, 8, 12]) in this area report that plaque morphology is also an important ultrasound marker that positively correlates with symptoms and identifies symptomatic plaques from asymptomatic ones.

2 Problem Formulation and Data

In this study the problem is formulated in two steps i) extract a set of features, coding textual information, that will be used by the classifier; ii) study the discriminative power of such features, in a real setting using both classical statistics and machine learning feature selection, in the classification problem of deciding if a carotid lesion (plaque) is associated (or not) to neurological events (symptoms).

The available dataset contains 146 ultrasound b-mode images of real carotid arteries with plaques, from a cross-sectional study of 99 patients (75 males and 24 females with a mean age of 68 years (41-88)) acquired at Cardiovascular Institute of Lisbon and Department of Vascular Surgery, Hospital de Santa Maria, Lisbon. The ground truth of this database is $N = 102$ (70%) and $P = 44$ (30%).

2.1 Monogenic decomposition

The monogenic filters (MF) [3], an extension to 2D proposed by [5] allows for an analytic decomposition of a 2D image into a) amplitude (A), b) phase angle ($\phi$) and iii) phase orientation ($\theta$), while preserving symmetry, energy, allpass transfer function, and orthogonality.

For an observed image $f(x)$, with $x \in \mathbb{R}^2$, Felsberg defines the three-component monogenic function as:

$$f_M(x) = [f(x), R_x(f)(x), R_y(f)(x)] = (f, R_x * f, R_y * f)$$

where * stands for convolution and $R$ is the Riesz transform with the following spatial representation

$$(R_x(x), R_y(x)) = \left(\frac{x_1}{2\pi |x|^2}, \frac{x_2}{2\pi |x|^2}\right), \quad x = (x_1, x_2) \in \mathbb{R}^2$$

The amplitude, orientation and phase are respectively calculated by:

$$A(x) = |f_M(x)| = \sqrt{R_x^2\{f\}(x) + R_y^2\{f\}(x)}$$

$$\theta(x) = \arctan\frac{R_y\{f\}(x)}{R_x\{f\}(x)}, \quad \theta \in [0, \pi)$$

$$\phi(x) = \arctan2\left(\sqrt{R_x^2\{f\}(x)R_y^2\{f\}(x)}, f(x)\right), \quad \varphi \in [0, \pi)$$

To fully describe a signal, usually multi scale monogenic representation is required, hence band-pass filtered versions of the input signal are used to achieve multi-resolution analysis.

![Figure 1: Example decomposition of an speckle image with plaque ROI into the monogenic signals for the first level os detail](image)

2.2 Feature Extraction

Here we extract the following features from the three monogenic signals: energy, mean and average parameters of a 2D-order 1 auto regressive (AR) model and study their discriminative power. These feature are calculated as:

$$E_y = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{i,j}^2$$

$$\mu_y = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{i,j}$$

$$y_{i,j} = \sum_{k,l} [a_{k,l} y_{i-k,l-j} + w_{i,j}, (k,l) \in (0,1); (1,0); (1,1)]$$

![Image](image)

RecPad 2013 Page 136 of 147
where $a_{k,j}$ are the AR coefficients to be estimated, $e_{i,j}$ are the residues and $y$ is the decomposed image obtained as follows: The US Bmode image is first normalized to $[0,255]$, then decompressed according to [10] to obtain a closer estimate image of the original Radio Frequency image. Then it is decomposed into its speckle (see Fig. 1) and despeckle fields according with [10], where the first contains most of the relevant texture information that will be used. The speckle image is separated into three levels of textual detail by using three band-pass filters. Each of these levels are decomposed in their monogenic signals (4) resulting in 9 images. Energy and mean features are calculated within the plaque ROI (segmented by an expert clinician), while the AR parameters are averaged over 50 random windows sampled from within the ROI.

### 3 Results

It is the purpose of this paper to investigate the application of monogenic decomposition in identifying symptomatic v.s. asymptomatic. Here we use the Mann-Whitney classical statistics approach to that performs a two-sided rank sum test of the null hypothesis that feature values in symptomatic and asymptomatic populations are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians.

#### 3.1 Feature Selection

Furthermore we investigate the usefulness of the presented features in a machine learning approach, by using Pudil’s Sequential Floating Forward Feature Selection (SFFS) [11]. Although computationally more expensive that standard forward feature selection approaches, this method avoids the nesting effect by allowing the removal of features selected in previous steps of the algorithm.

Along with the proposed features (see section 2.2) SFFS received a dataset with a series of other 116 relevant features ([9]) including morphological features extracted by the clinical, histogram features, Rayleigh Mixture Model features and textural features from grey level co-occurrence matrices (GLCM) and Wavelets.

For comparison purposes, notice that features (5) are also extracted from the standard Haar wavelet decomposition with 4 levels.

### 4 Discussion and Conclusions

Using US imaging and it’s speckle decomposition, we investigate the carotid plaque symptomatic predictive power of features extracted from monogenic signal decomposition. In a relevant database of 146 plaques, several monogenic features show high correlation with symptomatic v.s. asymptomatic classes (see table 1). Special notice for the AR model features, used to describe the spatial correlation between pixels.

Incorporating the proposed features into a machine learning feature selection allowed to compare the monogenic framework against other features. Interesting to note that although wavelets and monogenic features should encode similar textural information, the SFFS method clearly selected the later showing it’s better performance for classification of plaque symptomatic status.

This motivates the pursuit of the monogenic framework for more in-depth textual analysis of the carotid atherosclerotic disease.

### References


Voice Type Discovery

Mário Amado Alves
amadoalves@fe.up.pt
Vitor Almeida
valmeida@fe.up.pt
Ricardo Sousa
sousa.ricardo@fe.up.pt
Aníbal J. S. Ferreira
ajf@fe.up.pt
Sérgio I. Lopes
sil@fe.up.pt

Abstract
We address the problem of automatically recognizing the voice type, or tessitura, of singers. Such procedures have applications to singing learning, training, and entertainment. The main non-trivial component of our approach is the robust extraction of the fundamental frequencies present in a sample of the singing voice.

1 Introduction
We address the problem of automatically recognizing the voice type, or tessitura, of singers. This is an exploratory study. We consider the classical definition of voice types: bass, baritone, tenor, alto, mezzo-soprano, soprano [1,2].

The automatic recognition of the voice type of a singer has applications to singing learning, training and entertainment. For example, upon voice type discovery, a karaoke-style game may automatically adapt its pitch-matching algorithm to better suit the voice type of the singers, thus allowing players with different voice types to compete fairly on songs that would otherwise favour one pitch range over the other. In fact, such a game, called SingingBattle, designed for deployment at Casa da Música in Porto, is currently under development by ourselves in the context of the ARTTS project (gnomo.fe.up.pt/~voicestudies).

2 Algorithms
We have developed software libraries and computer applications to let the users discover their voice type, by singing, to the connected microphone, the passage described on Figure 1, which is our representation of the procedure as prescribed in [2], namely:

- Begin singing a note that is somewhere in your lower middle range.
- Sing a chromatic scale downwards in pitch. Write down the lowest note that you are able to vocally produce. Then, beginning at a comfortable upper-middle note, begin singing a chromatic scale upwards in pitch.
- Write down the highest note that you are able to sing.

Figure 1: Singing passage designed to discover the voice type. The cross note-heads represent approximate, imprecise pitches.

The overall algorithm, depicted in Figure 2, takes the sound of the singer performing the voice discovery passage (Figure 1), extracts the fundamental frequencies involved, then finally classifies the voice type based on an analysis of these frequencies. The details of each component are described in the next subsections.

2.1 Pitch and frequency
We use the mathematical notion of musical pitch, and the widely accepted pitch notation described in [3]. Namely, we call pitch to the musical note related to acoustic pitch and frequency as follows. Pitch classes are the notes represented by the letters and/or respective number of semitones in Table 1.

<table>
<thead>
<tr>
<th>C</th>
<th>C#</th>
<th>D</th>
<th>D#</th>
<th>E</th>
<th>F</th>
<th>F#</th>
<th>G</th>
<th>G#</th>
<th>A</th>
<th>A#</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 1: Pitch class symbols and semitones.

The pitch classes divide the octave into twelve equal logarithmic steps of $2^{1/12}$, called semitones. The octave is the interval between a frequency $f$ and its double $2f$. The names octave, semitone and the note letters have merely historical significance.

By international convention, octaves are numbered, and each octave starts at a C. A pitch class $K$ at octave $N$ is notated $K_N$ or $K_0$. To clarify: a pitch is a frequency. Also, by convention we have:

$$A_4 = 440Hz.$$  

To find the frequency of any note we work from (1) and the above rules. To simplify, we work with *semitone indices* across all octaves defined as

$$i(K_N) = 12(N - N_{ref}) + i(K) - i(K_{ref})$$  

where $i(K)$ is the function defined by Table 2. Then the frequency in Hertz of a pitch index is given by

$$f(i) = f_{ref} 2^{i/12}$$

where $s$ is the semitone interval $s = 2^{1/12} \approx 1.06$. The reference values must be consistent across (2) and (3). Normally we choose reference values tuned (no pun intended) for MIDI note numbers, for interability with MIDI components, namely $N_{ref} = -1, K_{ref} = C, f_{ref} = 440Hz, i_{ref} = 69$.

To derive the pitch of a given frequency we use approximation. Namely, the pitch index $i$ of a given frequency $f$ is the one such that $f(i)$ is the closest to $f$, on the logarithmic scale.

Figure 2: Overall algorithm

2.2 $f_0$ Extracton
Our algorithm is based on an analysis of the fundamental frequencies $f_0$ produced by the singer. The extraction of the $f_0$ values of the singing voice is a non-trivial matter [4,5]. We use a combination of techniques including peak-picking for harmonic location [5], interpolation of the spectral magnitude for $f_0$ estimation precision [4], and harmonic spectral magnitude based weighing [5]. See these references for detailed descriptions.

This same method of $f_0$ extraction has proven successful also in other applications, such as real-time audio compression [5], real-time pitch modification, and real-time pitch tracking for karaoke-like gaming, also under development at our laboratory (gnomo.fe.up.pt/~voicestudies).
2.3 Voice type classification

For the final analysis of the extracted pitches, we assume the production has respected the score in Figure 1. With this assumption a simple approach based on the mean and median of the pitches is attainable. Basically, we match the sample’s mean/median pitch with the reference’s mean/median pitch derived from the ranges described in Table 2. Namely, we select the voice type which mean/median pitch is the closest to the sample’s mean/median pitch (in any direction).

<table>
<thead>
<tr>
<th>Voice type designation</th>
<th>Pitch range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bass</td>
<td>D2 – E4</td>
</tr>
<tr>
<td>Baritone</td>
<td>F2 – G4</td>
</tr>
<tr>
<td>Tenor</td>
<td>A2 – D5</td>
</tr>
<tr>
<td>Alto</td>
<td>B3 – E5</td>
</tr>
<tr>
<td>Mezzo-soprano</td>
<td>G3 – A5</td>
</tr>
<tr>
<td>Soprano</td>
<td>B3 – C6</td>
</tr>
</tbody>
</table>

Table 2: Common designations and pitch ranges of the six main classical voice types, adapted from [1]. For the pitch notation see section 2.1

Figures 3-5 contain screenshots of the voice discovery function at play in the SingingBattle application under development. When a player presses the “Discover!” button on their “Mic” panel (Figure 3), the voice discovery widget shows up (Figure 4), consisting of a pitch ruler of sensible range (C2-C7). Upon performance, by the singer, of the voice discovery passage, the acquired pitches are plotted in real-time upon the ruler as white dots. Eventually the voice type classification takes place and is manifested by a label in the widget and by the selection of the corresponding checkbox on the players “Mic” control widget (Figure 5). This design is still under construction.

3 Exploratory results and things to try

Systematic evaluation has not been done yet. This is an exploratory study. Extensive ad hoc testing has shown that the above method works well more often than not. But there are occasional misclassifications. A number of possible improvements is envisaged.

The current method requires the disciplined action of the singer to perform the prescribed passage. A possible improvement in the direction of a more natural, less contrived interaction with the singer would be to allow any singing to be produced. Actually the simple mean/median pitch method already used seems applicable to such unconstrained sample, because it does not look at the sequential aspect. That is, all other things being equal, it seems the prescribed passage is not required after all.

Unfortunately, it might not be the case that all other things can be maintained equal. Another possible improvement is to make the procedure more robust to intervening noise. The method as is includes any \( f_0 \) that may have its source in extraneous elements to the singer e.g. other voices than the singer, office noise, etc. Such noises are likely to produce outliers in the set of acquired \( f_0 \) values. We are currently investigating possible methods of identifying such outliers. The methods we have envisaged so far resort to the sequential aspect of the input, namely the two phase curve form. That is, for these methods, we again require the disciplined input of Figure 1.

Yet another possibility is to use image pattern recognition methods for voice type classification [6]. The main idea is to generate a visual representation, i.e. a 2D image, containing the time-pitch data obtained in the voice type discovery exercise. It is expected that this spectrogram-like image be highly correlated with the image pattern that represents each voice type. This approach can effectively reduce the outliers impact on classification, thus improving the voice type classification rate.

4 Conclusion

We have developed \( f_0 \) extraction algorithms [4,5] which are adequate for the recognition of the voice type of singers. The algorithms are fast enough to be used in real time applications. We are developing computer software prototypes for interactive, real-time, voice type discovery (Figures 3, 4, 5).

Avenues of future research and development include the improvement in usability and performance of the voice type discovery functionality, and further computer applications that help to learn, train, or simply enjoy singing. A number of such applications are at various stages of development (gnomo.fe.up.pt/~voicestudies).

This work was supported by the Portuguese Foundation for Science and Technology, an agency of the Portuguese Ministry for Education and Science, under research project PTDC/SAU-BEB/14995/2008.

References


Sialolith metrics computed from microtomography data

Pedro Nolasco
A.P. Alves de Matos
Paulo V. Coelho
Carla Coelho
António Mauricio
Manuel F.C. Pereira
Raúl C. Martins
João M.R. Sanches
Patricia A. Carvalho

CEMS, Instituto Superior Técnico, University of Lisbon
Department of Pathological Anatomy, Centro Hospitalar de Lisboa Central, Hospital Curry Cabral
Service of Maxillofacial Surgery, Centro Hospitalar de Lisboa Central, Hospital S. José
CERENA, Instituto Superior Técnico, University of Lisbon
IT, Department of Bioengineering, Instituto Superior Técnico, University of Lisbon
ICEMS, Department of Bioengineering Instituto Superior Técnico, University of Lisbon
ICEMS, Instituto Superior Técnico, University of Lisbon

Abstract

A quantitative characterization of the internal structure of sialoliths is necessary for a correct analysis of their mechanical response to shock waves, namely, to establish the relevant geometric features behind the relatively high fracture toughness of these calcified structures. The current work defines suitable parameters for the geometric characterization of salivary calculi, including their periodic growth patterns.

1 Introduction

The formation of salivary calculi, or sialoliths, in the intra or extra-glandular duct system occurs in 0.12 to 0.45 % of the general population [1]. Long-term obstruction, even in the absence of infection, leads to atrophy of the affected gland with concomitant ceasing of the secretory function and ultimately fibrosis [7]. Shock wave lithotripsy is a non-invasive therapeutic technique that uses shock waves to shatter salivary calculi into fragments small enough to be naturally excreted through the ducts of the salivary glands [2]. However, the method has a less than desirable success rate on parotid sialoliths (34 – 69 %) and on submandibular sialoliths (32 – 42 %) [8]. The high failure rate of the technique is poorly understood due to the absence of a general quantitative description of the sialoliths internal structure, required to correctly model their mechanical response to shock waves.

Although some structural diversity is present [3], sialoliths present in general a single nucleus surrounded by alternating organic and mineralized layers permeated by globular dispersions of organic matter [3]. The present work proposes a methodology for a high throughput quantitative characterization of the general morphology and periodicity of the internal structure of sialoliths.

2 Methods

2.1 Experimental

Digital radiographs have been acquired using a µCT SkyScan 1172 instrument with an X-ray cone incident on a rotating specimen [6]. Each sialolith was positioned with the major axis perpendicular to the holder plate at the center of the stage and was scanned with a source voltage and a current of, respectively, 100 kV and 100 mA, using downstream 0.5-mm aluminum filtration to minimize the “beam hardening” effect. The acquisition was performed by rotating the specimen over 360° with a step size of 0.7°. The pixel size varied between 5 µm and 15 µm according to the specimen size. The data were saved in attenuation coefficient units (mm⁻¹) as 16-bit TIFF files.

Slice reconstructions were performed with the NRecon 1.6.3 reconstruction software from SkyScan [5] using the ring artefacts reduction, beam-hardening correction, and misalignment compensation built-in routines. The tomograms were saved as 8-bit BMP files in grayscale units. Post-processing was performed with the CT-analysyer (CTan) software package from SkyScan [4], using basic image processing methods. Binary masks were generated from the reconstructed cross-sections after manual segmentation and background noise reduction by applying the CTan despeckle built-in function.

2.2 Metrics definition

The total volume, mean attenuation, aspect ratio, eccentricity and the periodicity of the growth pattern have been estimated for 2 sialoliths with algorithms developed in MATLAB®.

The 3D structure of the specimens was digitally reconstructed and the sialolith volume V and the mean intensity I of each specimen were calculated, respectively, from:

\[
V = \frac{1}{N} \sum_{i=1}^{N} n_i, \quad I = \frac{1}{N} \sum_{i=1}^{N} T_i
\]

where \(v_{\text{voxel}}\) is the volume of the voxel, \(N\) is the total number of points in the 3D reconstructed matrix and \(i\) is grayscale value of the \(n\)th voxel. The mean intensity \(I\) was converted to attenuation \(t\) using the NRecon conversion scale.

Radial and diameter intensity profiles, with origin at the nucleus of the specimen determined manually from the internal morphology, were taken in spherical coordinates (Figure 1 (a)) with a step size of 5°.

The aspect ratio \(A_r\) and the eccentricity \(Ecc\) of the specimens were calculated from the ratio between the maximum and minimum cardinality of the diameter profiles \(\{D\}_{k}\) and radial intensity profiles \(\{R\}_{l}\), respectively:

\[
A_r = \frac{\max \{D\}_{k}}{\min \{D\}_{k}} \quad \text{and} \quad Ecc = \frac{\max \{R\}_{l}}{\min \{R\}_{l}}
\]

Figure 1 – Diameter profiles obtained from the 3D reconstructions. Two-dimensional matrices were generated by concatenating the 1-D profiles.

The diameter profiles were then concatenated into columns (Figure 1 (b)) and compensated geometrically to the mean of \(\{D\}_{i}\) by linear interpolation (Figure 1 (c)). The resulting 2D matrix was then averaged along each row to obtain the mean intensity profile for each specimen, which was subsequently translated back into attenuation units using the NRecon conversion scale (Figures 2 (b) and 3 (b)).

The fundamental periods of the growth patterns determined from peaks with amplitude \(a\) above the standard deviation have been...
identified from the FFT amplitude spectrum of the mean attenuation profile (Figures 2 (c) and 3 (c)). The peak amplitude $a_p$ was found to decay with the spatial frequency $f$ according to:

$$a_p = \frac{K}{f}$$

(3)

The fitting parameter $K$ can be used to classify the periodicity of the sialolith growth pattern.

3 Results and discussion

Table 1 presents the values for volume, mean attenuation, eccentricity and aspect ratio of sialoliths 1 and 2. These results have been confirmed with the CTan software package. Table 2 presents the parameters inferred from Fourier analysis.

Table 1 – $S_1$ and $S_2$ metrics.

<table>
<thead>
<tr>
<th></th>
<th>Present algorithms</th>
<th>SkyScan built-in</th>
<th>Present algorithms</th>
<th>SkyScan built-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume, $V (10^3 m^3)$</td>
<td>85.01</td>
<td>85.07</td>
<td>88.65</td>
<td>88.813</td>
</tr>
<tr>
<td>Mean attenuation, $t (10^3 m^3)$</td>
<td>0.016</td>
<td>0.015</td>
<td>0.0045</td>
<td>0.043</td>
</tr>
<tr>
<td>Aspect ratio, $Ar$</td>
<td>2.19</td>
<td>0.015</td>
<td>2.29</td>
<td>-</td>
</tr>
<tr>
<td>Eccentricity, $Ecc$</td>
<td>2.45</td>
<td>-</td>
<td>2.78</td>
<td>-</td>
</tr>
</tbody>
</table>

$S_1$ presents an inferior number of fundamental periods and a lower fitting parameter than $S_2$ (Table 2), in agreement with the higher periodicity of the $S_1$ growth pattern (compare Figures 2 (a) and 3 (a)).

4 Final remarks

Shock wave lithotripsy is a non-invasive therapeutic technique that presents a low efficiency on treating sialolithiasis. A comprehensive analysis of the effect of shockwaves on the calculi structure is still lacking, and this type of study demands sound quantitative information about the sialoliths structure.

This work proposes a method for a high throughput quantitative characterization of some geometric parameters of sialoliths, such as: volume, aspect ratio, eccentricity, mean attenuation and periodicity of the growth patterns.

Further studies will be carried out to better understand the effect of shockwaves on the calculi structure.

References


Figure 2 – Characterization stages of sialolith $S_1$. a) Median cross-section. b) Mean 1D intensity profile and c) Amplitude spectrum of the mean intensity profile. Dashed line: $Kf$, dotted line: standard deviation of fundamental amplitudes.

Figure 3 – Characterization stages of sialolith $S_2$. a) Median cross-section. b) Average 1D intensity profile and c) Amplitude spectrum of the mean intensity profile. Dashed line: $Kf$, horizontal line: standard deviation of fundamental amplitudes.
Automatic gesture segmentation based on a predictive event segmentation approach

Sofija Spasojevic
http://www.pupin.rs/RnDProfile/spasojevic.html

Rodrigo Ventura
http://users.isr.ist.utl.pt/~yoda

Mihailo Pupin Institute, Robotics department
Faculty of Electrical Engineering, University of Belgrade
Institute for Systems and Robotics
Instituto Superior Técnico, Lisbon

Abstract
This paper presents preliminary work related to one approach for automatic gesture segmentation. We take a predictive event segmentation approach, according to which events are detected once sensor data departs significantly from an adaptive model-based predictor. During gesture execution, 3D positions of characteristic human joints are collected using a Kinect device. Data set consists of joints from interest, whose trajectories are modelled as Gaussian processes. Based on this data set, a Gaussian process based predictor is adapted and used to detect transitions between gestures. The preliminary results over the collected dataset are encouraging and illustrate the outcome of the approach.

1 Introduction
Recent years, research in area of human gesture recognition have become very intensive and number of publications related to this area have increased significantly. Gesture recognition systems have significant applications in many different fields such as virtual and augmented reality [1], industrial process control [2], physical rehabilitation [3], human–robot interaction [4], computer games [5] etc.

Each gesture recognition system is a complex structure which contains from several functional units. First step in every gesture recognition process is gesture acquisition. A lot of techniques for gesture acquisition, based on different type of sensors have been developed. In general, all sensors used for collecting information about human movement can be categorized in two groups – body and visual sensors. Body sensors are placed on the body parts and provide information directly each time the movement occurs. Some examples of most used body sensors are accelerometers, data gloves and special body suits that using optical or electromechanical tracker technologies. Visual sensors refers to camera systems used for recording gesture sequences, followed by image processing techniques in order to characterize each gesture with a set of visual features. Some gesture acquisition techniques include both type of sensors in order to obtain more accurate information. Attaching sensors or body markers on the body, as well as wearing special suits many people find uncomfortable. On the other hand, visual sensors can be very expensive and their performance depends a lot of illumination and background conditions. Considering above mentioned items, we decided to use Kinect device for gesture acquisition process.

Kinect is the new generation low-cost device developed by Microsoft, which consists of variable-resolution RGB camera, 3D depth sensor (infrared projector combined with infrared camera) and broadband microphone. Kinect has a corresponding user interface and compatibility to work with different software packages. It also has some embedded algorithms, one of which relates to skeleton detection and collecting 3D joints position. This algorithm presents a base for gesture acquisition process used in this work.

Another important topic when dealing with gestures is gesture segmentation, which is the pre-phase of applying gesture recognition algorithm. This phase is very important, considering the influence that inaccurate segmentation can have on process of gesture recognition. There are a lot of techniques applied in gesture segmentation problems.

Technique based on simple sliding window combined with simple moving average filter is used in [6]. Author defines content of each gesture in the following form: starting static posture, dynamic gesture part and ending static posture. In addition, to obtain more robust segmentation, author observe also the length of each analyzed sequence to eliminate appearance of static part into dynamic part of gesture. In [7] authors developed algorithm for segmentation of dance sequences. This algorithm, called Hierarchical Activity Segmentation, is based on division of human body onto hierarchically dependent structures. They take into account relevant motion parameters for body segments

(1)

where \(X\) is the known function values of the training cases and \(X_i\) is a set of function values corresponding to the test set inputs, \(\mu = m(x_i), i=1,\ldots,n\) for the training means and analogously for the test means \(\mu;\) for the covariance we use \(\Sigma\) for training set covariances, \(\Sigma_e\) for training-test set covariances and \(\Sigma_c\) for test set covariances. Since the values for the training set are known, we should determine conditional probability distribution of \(X_i\) given \(X\), which is expressed as:

(2)

This is the posterior distribution for a set of test cases. Very important item when dealing with Gaussian predictive models is process of hyperparameters estimation. Depending on the form of training and testing set samples, number of hyperparameters that define mean and covariance function can vary. Estimation of hyperparameters is achieved by maximization of the log-likelihood function.

Acknowledgement: This work was partially funded by bilateral project COLBAR, between Instituto Superior Técnico, Lisbon, Portugal and Mihailo Pupin Institute, Belgrade, Serbia, and FCT [PEst-OE/EEI/LA0009/2013], III44008 and TR35003.

Figure 1: Skeleton tracking (a) and joints whose 3D positions are collected (b) by Kinect

(segmental force, kinetic energy and momentum) that characterize motion in the levels of defined hierarchy. In [8] the authors took a dynamical system approach for dynamic system identification, however that approach did not account for sensor noise. In this paper, we study an approach using Gaussian processes as machine learning method [9] that provides information about both, value and uncertainty. In addition, this method has shown good properties related to complexity model and processing time.

2 Algorithm for gesture segmentation

As previously mentioned, process of gesture acquisition is realized using Kinect device. Some functions from OpenNI software package designed for work with 3D sensors were used in order of collecting information from Kinect. During skeleton tracking, 3D positions of 15 joints (Fig. 1) for each frame are collected. In this preliminary work we concentrate on segmentation of upper body gestures, concrete gestures performed by arm (Fig. 2). Accordingly, only elbow and hand joints are important for further analysis. After normalization of coordinates with respect to torso, data set is configured.

We modelled trajectories of elbow and hand joint positions as Gaussian processes and formed three predictive Gaussian prediction model (each model for one coordinate). If we want to make predictions for following values from existing set of normally distributed variables, we can define joint distribution as:

\[
\begin{bmatrix}
X^t \\
X^\ast
\end{bmatrix}
\sim N
\left(
\begin{bmatrix}
\mu \\
\mu^\ast
\end{bmatrix},
\begin{bmatrix}
\Sigma & \Sigma_e \\
\Sigma_e^\ast & \Sigma_c
\end{bmatrix}
\right)
\]

where \(X\) is the known function values of the training cases and \(X^\ast\) is a set of function values corresponding to the test set inputs, \(\mu = m(x), i=1,\ldots,n\) for the training means and analogously for the test means \(\mu;\) for the covariance we use \(\Sigma\) for training set covariances, \(\Sigma_e\) for training-test set covariances and \(\Sigma_c\) for test set covariances. Since the values for the training set are known, we should determine conditional probability distribution of \(X^\ast\) given \(X\), which is expressed as:

\[
X_i^\ast | X \sim N(\mu_i + \Sigma_e^{-1}(X - \mu), \Sigma_e^{-1} + \Sigma_c^{-1})
\]

This is the posterior distribution for a set of test cases.
The optimization requires the computation of the derivative of log-likelihood function with respect to each of the parameters. Let \( n \) be the number of frames in gesture sequence and \( x_i \) value of \( x \)-coordinate in \( i \)-th frame. Training set consists of samples that are organized as five dimensional vectors:

\[
X = ([x_1, \ldots, x_5]; [x_6, \ldots, x_8]; \ldots; [x_{n-4}, \ldots, x_n])
\]

(3)

Testing set contains from scalar samples that represent first value that follow appropriate training sample:

\[
X_i = (x_{i,1}; x_{i,2}; \ldots; x_{i,5})
\]

(4)

Analogously in case of the training and testing set for \( y \) coordinate, \( Y \) and \( Y \). Given this data set, corresponding mean functions of Gaussian models have per five, and covariance functions per two free parameters, that are determined in the process of hyperparameters optimization.

Predictive models are defined using training and testing set, obtained hyperparameters and selection of appropriate inference method. Models are formed for \( x \) and \( y \) trajectories of hand joints, since the gestures are performed in \( x-y \) plane. Observations of only one coordinate wouldn’t be enough, given that these maybe don’t include information necessary for description of all gestures. Values of \( z \)-coordinate in this case didn’t give any contribution to final result, therefore they are not taken into account.

Determining of predictive values starts from the beginning of sequence and during prediction progress training set increases. Error of prediction in form of difference between real \( (\hat{x}, \hat{y}) \) and Mahalanobis distance (5) are calculated each step. When Mahalanobis distance for several successive moments increases significantly, those small intervals are marked as event, ie potential start or end of gesture. After detecting this discontinuity, training set resets and starts to form again from first next sample onwards. Prediction process continues at the same way until end of gesture sequence.

\[
MD = \sqrt{[x-\hat{x} \ y-\hat{y} \ \sigma_x \ 0 \ \sigma_y] [\begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}^{-1} \begin{bmatrix} x-\hat{x} \\ y-\hat{y} \end{bmatrix}]}
\]

(5)

where \( \sigma_x \) and \( \sigma_y \) are predictive variances for first and second Gaussian predictive model, respectively.

### 3 Results

One of analysed gesture sequences (Fig. 3) contains from three different gestures (Fig. 2), each of them performed twice. Trajectories of \( x \) and \( y \) coordinate of hand wrist for this gest sequence are shown on the Fig. 3 (top) together with predicted values. Variations of Mahalanobis distance through the time are also shown on Fig. 3 (bottom). Red points on this graph indicate an event. It can be seen that Mahalanobis distance increase every time that one of coordinates suddenly drops or increases value and form peaks on these small intervals. Note that red points are positioned at the beginning or end of the gestures.

### 4 Conclusion and future work

In this preliminary study we have analysed one approach for gesture segmentation based on predictive Gaussian model. This approach has shown very good performance according to the criteria of model complexity and time necessary for algorithm processing. For few analysed gesture sequences until now, our method has shown excellent results in the sense of correct detection of significant changes in gesture performing. Based on these results, gesture segmentation can be performed directly.

Future work will be oriented to improvement of this method and generalization in case of larger and more diverse gesture sequences.

### References


Simultaneous Model Estimation, Denoising, and Noise Decomposition

Manya V. Afonso
mafeson@isr.ist.utl.pt
Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisbon, Portugal

Joaõ M. R. Sanches
jmrs@isr.ist.utl.pt

Abstract

In some situations, the noise contaminating an image may contain useful information, thereby necessitating the separation of the noise field along with estimating the image. When the algebraic observation model is known, the noise component can be estimated in a straightforward manner after denoising. However, when the model is unknown or if some data are missing, we need to make some assumptions about the nature of noise field. In this paper, we propose a method for simultaneously determining the algebraic model, estimating the image, and separating the noise field.

We use a polynomial representation to relate the image and noise with the observed image, and estimate the polynomial coefficients, the image, and noise component iteratively through alternating minimization.

1 Introduction

The noise contaminating a signal or image can sometimes hold important information such as texture, that may be useful, especially if it is signal dependent noise. For example, ultrasound (US) images are corrupted with multiplicative speckle noise which some authors argue is not truly noise because it is dependent of the intrinsic acoustic properties of tissues, and therefore, its texture often carries useful information for characterization of tissues and/or pathologies (see [11] and references therein). Noise field estimation has been reported for US speckle [11], and rician noise in MRI [1, 4].

Denoising and reconstruction methods focus on estimating the desired noiseless image \( x \), given the noisy observed image \( y \), without taking into account the noise field \( \eta \) as a signal component to be estimated. Signal and image decomposition methods [14], [12], [13] separate the smooth (cartoon) component and texture component of an image, but discarding the noise component.

When the size of the noiseless and observed images is the same, i.e., there is no loss of data, the observed field can be computed knowing the observation \( y \) and the estimate of the noiseless image \( \hat{x} \), if the algebraic model for the image observation is known. When the algebraic model is not known or it is known but there are some data missing, it is not possible to estimate the field through the inverse operation of the algebraic model. In the case of partial observations, a suitable regularizer needs to be assumed for \( \eta \), intuitively, one that discourages sparsity. The choice of this regularizer would also naturally depend on an assumption about the statistical distribution of the noise.

1.1 Contributions

In this paper, we propose a method for denoising and estimating the noise field, given the noisy observed image without missing pixels but without knowledge of the algebraic model. We treat the noise as one of the signal components to be estimated and as in [12],[13], we use a polynomial representation to relate the image and noise components, and iteratively estimate the image, noise field, and polynomial coefficients through alternating minimization.

2 Proposed Method

Without knowledge of the algebraic observation model, we use a polynomial representation to relate the image \( x \) and noise \( \eta \):

\[
 y = f(x, \eta, a) = a_0 + a_1 x + a_2 \eta + a_3 x \eta,
\]

where the vector of coefficients \( a = [a_0, a_1, a_2, a_3]^T \) is also a variable in the estimation process. The order of the polynomial will of course have to be different to account for higher order terms, such as the squared amplitudes, in the MRI acquisition model. The order used in (1) is sufficient to account for the additive \( y = x + \eta \), (a = [0, 1, 1, 0]^T) and multiplicative \( y = x \eta \), (a = [0, 0, 0, 1]^T) models.

We define the discrepancy term for the model from (1) as the squared \( \ell_2 \)-norm of the difference,

\[
 J(y, \hat{x}, \hat{\eta}, \hat{a}) = 0.5 ||y - f(\hat{x}, \hat{\eta}, \hat{a})||_2^2,
\]

where \( \hat{x}, \hat{\eta}, \hat{a} \) are the respective estimated values. We formulate the problem of estimating \( (x, \eta) \) as a convex optimization problem, with a total variation (TV) [10] regularizer on \( x \) to promote piece-wise smoothness, and a quadratic norm regularizer on \( \eta \) to lead to a non-sparse estimate [5], as follows,

\[
 \min_{x, \eta, a} J(x, \eta, a) + \frac{\lambda_\eta}{2} \text{TV}(x) + \frac{\lambda_x}{2} ||\eta||_2^2,
\]

where \( \lambda_x, \lambda_\eta > 0 \) are the regularization parameters and \( \text{TV}(\cdot) \) is the isotropic TV function.

We solve (3) using the Augmented Lagrangian(AL)/Alternating Direction Method of Multipliers (ADMM) method [6], [15], [7], because of its mathematical elegance and computational simplicity. This leads to a Gauss-Seidel process in which at each iteration we estimate each variable from \( x, \eta, a \), keeping the other two fixed. Applying AL/ADMM to (3) leads to the following iterative process where \( u \) and \( \theta \) are the respective auxiliary variables for \( x \) and \( \eta \):

\[
 u^{(t+1)} = \arg \min_u \frac{\mu_x}{2} ||x^{(t)} - u - d_u^{(t)}||^2 + \frac{\lambda_x}{2} \text{TV}(u),
\]

\[
 x^{(t+1)} = \arg \min_x J(y, x, \eta^{(t)}, a^{(t)}) + \frac{\mu_x}{2} ||x - u^{(t)} - d_u^{(t)}||^2,
\]

\[
 \theta^{(t+1)} = \arg \min_\theta J(y, x, \eta^{(t)}, a^{(t)}) + \frac{\mu_\theta}{2} ||\eta - \theta - d_\eta^{(t)}||^2 + \frac{\lambda_\theta}{2} ||\theta||_2^2,
\]

\[
 \eta^{(t+1)} = \arg \min_\eta J(y, x^{(t)}, \eta, a^{(t)}) + \frac{\mu_\eta}{2} ||\eta - \theta^{(t+1)} - d_\eta^{(t)}||^2,
\]

\[
 d_u^{(t+1)} = d_u^{(t)} - x^{(t+1)} + u^{(t+1)},
\]

\[
 d_\eta^{(t+1)} = d_\eta^{(t)} - \eta^{(t+1)} + \theta^{(t+1)}.
\]

The variables \( d_u, d_\eta \) are the so-called Bregman weight vectors [15], and \( \mu_x, \mu_\eta > 0 \) are the weight terms for the constrained optimization (see [2] for details). The step (4) is computed using a few iterations of Cambronne’s algorithm [3]. Steps (6), (5), and (7) are least squares minimization problems and have linear closed form solutions.

At each iteration, the vector of coefficients \( a \) is updated using a polynomial fitting in two variables [9], [8], using the values (\( x^{(t+1)}, \eta^{(t+1)} \)).

3 Experimental Results

The method was implemented on Matlab running on an Ubuntu Linux laptop with an Intel i5. Fig 1(a) shows the size 128 x 128 Shepp-Logan phantom, which we use for our synthetic example in which we generate and add Gaussian noise, and demonstrate the estimation of the image, noise component, and polynomial coefficients. In this case, the synthetic noise we add to our image is considered the ground truth for noise. The observed noisy image after adding noise corresponding to a Signal to Noise Ratio (SNR) of 0 dB is shown in fig 1(b).

The estimate of the image was initialized to the noisy image, the estimate of the noise was initialized to a randomly generated noise vector with the same variance, and the estimates of the polynomial coefficients were all initialized to one. Figure 1(c) shows the estimate of the image after 500 iterations. The profile along the diagonal of the original image and noisy image are shown in fig 2(a). Figures 2(b) and 2(c) show the corresponding diagonal profiles of the original, i.e., ground truth, of the image and noise components, respectively, along with the estimated component.
proposed method. We see from the estimated image and diagonal profiles of the image and noise components, that the proposed method estimates the image and noise components closely, without prior knowledge of the algebraic model. From fig 3, we see that the final estimated polynomial coefficients are close to the true values.

Figure 1: (a) Original Shepp Logan phantom image, (b) noisy image with additive Gaussian noise (SNR: 0 dB), (c) final estimate using the proposed method.

Figure 2: Diagonal profiles of: (a) noisy image, (b) estimated image, (c) estimated noise.

Figure 3: True and estimated values of the polynomial coefficients describing the algebraic model.

Figure 4: Evolution of the MSE over time for the: (a) image, (b) noise field.

4 Conclusions

We have proposed a method for denoising and noise separation without knowledge of the algebraic model relating the image, noise, and observed image. The method models the interaction between the image and noise using a polynomial approximation, and also iteratively estimates the polynomial coefficients. Preliminary results for the case of additive and Gaussian noise show that the proposed method succeeds in estimating accurately the image, noise field, and model coefficients. The problem of noise estimation with missing data will be addressed in a future paper.

Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia (FCT), Portuguese Ministry of Science and Higher Education, through a Post-doctoral fellowship (contract no. SFRH/BPD/79011/2011) and FCT project (PEst-OE/EEI/LA0009/2013).

References

Abstract

In this paper we present a method which intends to estimate the position and orientation of a Ultra-Sound probe. The detection is based on the RuneTag-43 marker, with a ring-refitting and an iterative ring decoding approach. The presented results are based on the occlusion robustness of the marker.

1 Introduction

The 3D reconstruction of the carotid bifurcation is an important tool to characterize, evaluate and diagnose cases of atherosclerosis [4]. However, most of the 3D reconstruction systems are expensive and impractical. In this paper, we propose a method that aims to be simple, practical and inexpensive in the clinical use.

A free-hand Ultra-Sound (US) allows us to obtain the plaque sections. These sections, conciliated with their position and orientation, make the 3D reconstruction possible. The proposed solution is based on the algorithm and the marker presented in [2]. The idea is to locate the marker on the US probe, which is in the Field-of-View (FOV), in order to estimate the position and orientation of the section measured. The marker should be robust to lighting variations and resistant to occlusions, that can occur during the US exam. The marker must constrain the least possible, the proper execution of the medical examination. Regarding the detection algorithm, it intends to focus on the accuracy of the results, rather than the data processing speed. Actually, it is not quite likely that clinical doctor can afford the time required to do a 3D reconstruction during the exam and, in many cases, it is not essential for a proper analysis of the patient. Our method is a work in progress which intends to be off-line so that it can focus, as much as possible, on accuracy and ensure more realistic results.

2 Method

The method uses only one camera to estimate the position and orientation of a known object, the marker (Fig: 1.a).

2.1 Image Data Ellipse Detection and Conics

In general, the result of the projection of a circle in the image plane is an ellipse. Thus, each of the dots (Fig:1.a) that are present in our marker, will be "transformed" into ellipses on the image. Keeping this in mind, we aim to identify which of the ellipses are, in fact, dots that belong to our marker and eliminate the ones that are not (False-Positives, FP).

The Oblique Elliptical Cone (OEC) is a particular case of (2), where the conic section corresponds to a circle rather than an ellipse. From this point on, our goal is to find the relationship that transforms an elliptic section into a circular section. All points of the OEC that belong to \( Q_c \) can be described according to,

\[
P^T Q P = 0 \quad \text{with} \quad Q = \begin{bmatrix} A & B & D \\ B & C & F \\ D & F & E \end{bmatrix}.
\]

The Oblique Circular Cone (OCC) is a particular case of (2), where the conic section corresponds to a circle rather than an ellipse. From this point on, our goal is to find the relationship that transforms an elliptic section into a circular section. All points of the OCC that belong to \( Q_c \) can be described according to,

\[
P^T Q P = 0 \quad \text{with} \quad Q = \begin{bmatrix} 1 & 0 & -\frac{r}{s} \\ 0 & 0 & -\frac{g}{s} \\ -\frac{r}{s} & -\frac{g}{s} & -\frac{r^2+s^2}{s} \end{bmatrix}.
\]

Where \( Q_c \) represents the conic section of a circle with radius \( r \) and center \((x_0, y_0)\) in the plane \( z = 0 \).

Assuming that the origin of the OCC is the optical center and its Z-axis is given by the normal of the circle, since both oblique conic frames are located at the same origin, there exists a rotation \( R_c \) that relates the two coordinate systems

\[
P = R_c P_c.
\]

Our aim is to calculate that rotation, which is the same for all ellipses belonging to the marker, at less than a parameter (\( \alpha \)). Relating (2) with (3) and since (4),

\[
P^T Q_c P_c = w P^T (R_c^T Q R_c) P_c, \forall w \neq 0 \Rightarrow Q_c = w R_c^T Q R_c, \forall w \neq 0.
\]

2.2 Transformation Between Frames

In order to calculate \( R_c \), a Eigen Decomposition is applied to \( Q \). Replacing in (5),

\[
Q_c = w R_c^T V \Lambda^T R_c, \forall w \neq 0
\]

Where \( \Lambda \) and \( V \) are the eigenvectors and eigenvalues matrices of \( Q_c \) respectively.

The matrix \( R_c \) is, according to [3] given by,

\[
R_c = V \begin{bmatrix} g \cos \alpha & s_1 \cos \alpha & s_2 h \\ s_1 \sin \alpha & -s_1 \cos \alpha & 0 \\ s_2 h \cos \alpha & s_2 h \cos \alpha & -s_2 g \end{bmatrix}.
\]

In which \( g = \sqrt{\frac{2s_1}{s_2}} \) and \( h = \sqrt{\frac{2s_1}{s_2}} \) where \( s_i \) is the i-th eigenvalue, \( s_1 \) and \( s_2 \) are undetermined signs and \( \alpha \) is an arbitrary rotation angle among the normal of the dot (it was chosen \( \alpha = 0 \)). \((x_0, y_0, z_0)\) coordinates are determined by replacing (7) in (6) we conclude that,

\[
\begin{bmatrix} x_0 & y_0 & z_0 \end{bmatrix} = \begin{bmatrix} -s_2 z_1 \frac{1}{\sqrt{\lambda_2}} - s_1 z_2 \frac{1}{\sqrt{\lambda_1}} r & 0 & 3 s_3 \frac{1}{\sqrt{\lambda_3}} r \end{bmatrix}
\]

(8)

Since \( R_c \) depends on \( s_1 \) and \( s_2 \) and \( z_0 \) on \( s_3 \), there are 8 possible combinations between \( R_c \) and \( z_0 \). However, looking at the problem geometry, two of the three signs can be determined.

In the camera reference frame, a normal \( N \) and center \( C \), for each dot, are calculated and given by,

\[
C = R_c \begin{bmatrix} x_0 & y_0 & z_0 \end{bmatrix}^T \quad N = R_c \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T
\]

Since the marker is in the FOV, the following restrictions are applied,

\[
C \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T < 0 \quad N \cdot \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T > 0
\]

(10) from now on, only one of the three signs needs to be determined. Due to that, we have that two different \( R_c \) transformation matrices for each ellipse. Those are applied to the respective ellipse and two pairs of \( z_0 \) and \( N \) are obtained from each one. Those values are used to compare if, an ellipse belongs or not to the marker.
and dots (c) Refitted ring and pose estimation of a completely exposed marker (d) Same pose estimation as in (a) with 7 dots occluded.

Figure 1: (a) The marker, formed by 17 (dots) and 43 slots. The 17 dots form the ring; (b) Two elliptic rings that pass through the two accepted dots; (c) Refitted ring and pose estimation of a completely exposed marker (d) Same pose estimation as in (a) with 7 dots occluded.

Figure 2: 50 different data acquisitions (x axis), number of occlusions in each acquisition (black numbers) (a) the variations of \((x, y, z)\) coordinates relative to the mean position coordinates (blue, red, green) respectively, norm variations in (black). (b) Variations of \((\alpha, \beta, \gamma)\) orientation relative to the mean orientation angles (blue, red, green) respectively.

2.3 Best Fitted Dots and Ring Refitting

With two points on the circular plane it is possible to estimate the two circles that best fit these points. Thus, from the coordinates of a pair of dots, we can estimate the ring that fits the other dots. Our marker is flat and all the dots have the same and known radius \(r\). In the camera reference frame, each dot should present a normal vector that is parallel to the normal of the other dots and, in the circular frame, have the same value for \(z_0\). Thus, the parameters \(z_0\) and \(N\) are calculated for all ellipses. The chosen pair is the one that minimizes, \(e = e_0 + e_N\) with \(e_0 = |z_0 - z_o|\) and \(e_N = |1 - N_j \cdot N|^2\) for a pair \((i, j)\) of ellipses. From now on, with the chosen pair of dots, we can estimate the ring that contains all the other marker dots, since the radius of the ring is known and the rotation \(R\) that is applied to each conic dot, also affects the ring itself.

The conics \(Q_i\) and \(Q_j\) are calculated from a constrained average of the rotations \(R_{ij}\) and \(R_{ji}\) \(R_{i,j} = R_{ji}^{-1} R_{ij} = I\). Because both of the dots are located in the same \(z_0\) plane, the centers \((x_0, y_0)\) and \((x_0, y_0)\) are obtained assuming that \(z_0 = 0\). The radius of the ring is known so, from both dot centers in the circular plane, the two ring circles \((Q_{i,j})\) with \(l = 1, 2\) are calculated. Applying \(R_{i,j}^T\) to the conic ring circle \(Q_{i,j}\), two ring ellipses are obtained however (Fig.1.b), only one of them is really the ring that characterizes the marker and fulfills the requirements of (2) for each of the remaining dots. An acceptance criteria \(a\) is defined, so that, only the centers of the ellipses, that satisfies \(P^T Q_{i,j} P < a\), are accepted as marker points (dots). Thus, from the pair of rings that are projected in the image plane, it is accepted as marker final ring, the one that presents the largest number of acceptable points. The accepted marker dots allow us to refit the ring and \(R^t\) can be recalculated with (7).

2.4 Ring Decoding and Pose Estimation

The matching between the image dots \(P_i\) and the dots from the reference marker \(P_m\), Fig.1, is essential for the pose estimation step. The marker consists of 17 dots distributed in a circular pattern of 43 slots. Thus, starting at any dot in the circular frame, it is possible to create a marker whose distribution of dots is identical to the reference. By minimizing,

\[
e = \sum_{i=1}^{K} \min(||P_k - R - P_m||)\quad \text{with} \quad m = 1...M
\]

a rotation \(R\) that maximizes the matching of \(P_i\) and \(P_m\) is obtained where, \(K\) is the number of accepted dots, \(M\) the number of marker dots. Because \(K\) and \(M\) can be different, due to occlusions and miss-detections, the matching between the dots from both frames is not immediate. So the rotation matrix \(R\) is estimated iteratively,

\[
R = \begin{bmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

with \(\theta = \theta_{i,j} + \Delta \theta\) and \(0 \leq \theta \leq 360\). As each slot corresponds to \(\Delta \theta = 8.37^\circ\) of the ring, \(\Delta \theta = 6^\circ\) was chosen. From (11) and (12) is possible to do the labelling of each dot, according to the correspondence between the image and the reference marker dots.

Finally, since the pixel coordinates \((X, Y)\) of each dot in the image, and the coordinates \((x, y, z)\) of each dot in the own marker frame are known, as their correspondence, using the OpenCV function SolvePnP [1], the transformation (rotation + translation) that relates these coordinate systems, is calculated and the pose estimation is completed.

3 Results

The presented results focuses on the robustness evaluation of the marker position and orientation due to occlusions. The marker was placed at a distance of approximately 24.5 cm from the camera. Fifty acquisitions were made and, in all of them, the camera and the marker were fixed, but occlusions varied progressively from none to 7 occluded dots (Fig: 2.a and 2.b). Small variations in position and orientation should be observed. (Fig.2.a) shows the variation of the position and (Fig.2.b) the variation of the orientation, in relation to the average marker position and orientation. When the marker is completely exposed the variations are low \((\approx 0.1cm)\) which reveals sensitivity of the method for these cases. However, and particularly when semi-occlusion of a dot occur, there is a variation increase. The results show a maximum and sporadic variation of 0.853 cm, during a transition from 6 to 7 occlusions, probably due to the dot deformation that occurs from the occlusion, resulting in a change on the centroid position. The marker orientation \((\alpha, \beta, \gamma)\) Euler angles) shows, overall, small variations \((\{0^\circ, 3^\circ\})\) during acquisition. A particular peak in angle variation \((5.688^\circ, 5.645^\circ, 3.892^\circ)\) also happens when the number of occlusions changes from 6 to 7 dots.

4 Conclusion

We present a work in progress, but some conclusions can already be drawn. Results indicate the feasibility of the method even with strong marker occlusions. Future work will focus on making this analysis more realistic, defining metrics sensitive to changes like brightness, position, orientation and number of occlusions, enabling a more real and accurate position and orientation estimation.

5 Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia (PCT), Portuguese Ministry of Science and Higher Education, through the project (PEst-EEI/LA0009/2013).

References