RecPad 2012
Atas da XVIII Conferência Portuguesa de Reconhecimento de Padrões
Coimbra, Outubro 2012
Preface

The 18th Portuguese Conference on Pattern Recognition, RecPad 2012, took place at the Institute of Engineering of the Polytechnic Institute of Coimbra, on October 26, 2012. RecPad is an annual, one-day meeting, sponsored by the Portuguese Association for Pattern Recognition (APRP). In 2012 it was co-organized by the Institute of Engineering of the Polytechnic Institute of Coimbra. The RecPad 2012 was an opportunity to promote the collaboration between the Portuguese scientific community in the fields of Pattern Recognition, Image Analysis and Processing, Soft Computing, and related areas.

The response to the call for papers for RecPad 2012 was very positive and sixty nine papers were compiled for the proceedings. Each paper was reviewed by two reviewers. We would like to thank all the authors of the submitted papers for sharing their research activities. We also want to thank all members of the Technical Committee who spared their valuable time to review the submitted papers in a timely manner.

Finally, the local Organizing Committee is grateful to Carlos Gonzalez-Morcillo, Associate Professor of Computer Science at the University of Castilla-La Mancha in Spain, for having accepted our invitation to present a talk in the topic of “Indoor Navigation Infrastructure based on Augmented Reality”.

Coimbra, October 2012

Nuno Cid Martins
Verónica Vasconcelos
Fernando Lopes
Inácio Fonseca
Jorge Barbosa
Nuno Rodrigues
Simão Paredes
Organizing Committee

Nuno Cid Martins
Verónica Vasconcelos
Fernando Lopes
Inácio Fonseca
Jorge Barbosa
Nuno Rodrigues
Simão Paredes

Local Technical Support

Inês Duarte
Teresa Jorge
Paula Carvalho

Invited Speaker

Carlos Gonzalez-Morcillo

Technical Committee

António Pinheiro       Luís Roseiro
Armando Pinho          Manuel Crisóstomo
Carlos Pereira         Manya Afonso
David Afonso           Marco Gomes
Deolinda Rasteiro      Nuno Martins
Fernando Lopes         Nuno Rodrigues
Fernando Perdigão      Paulo Carvalho
Hans du Buf            Paulo Menezes
Hélder Araújo          Paulo Salgado
Inácio Fonseca         Pedro Assunção
Jaime Cardoso          Pedro Quelhas
João Sanches           Ricardo Ribeiro
Jorge Barbosa          Simão Paredes
Jorge Henriques        Verónica Vasconcelos
José Silva             Viriato Marques
Luís Cruz              Vítor Filipe
Luís F. Teixeira       Vítor Silva

Supported By

Instituto Superior de Engenharia de Coimbra
Associação Portuguesa de Reconhecimento de Padrões
Table of Contents

Invited Speaker
Gradient Flow Based Detection of Prominent Points on BCCT Depth-Map Images 3
Music Symbols Segmentation through Recognition 7
Denoising and Segmentation of the Second Heart Sound Using Matching Pursuit 9
A Relevance-based Linde-Buzo-Gray Approach for Feature Discretization 11
Supervised Feature Discretization by Mutual Information Maximization 13
Mobile Framework for Biosignal Acquisition and Remote Processing 15
Robotic 3D Ultrasound 17
Multimodal biometric recognition under unconstrained settings 19
A Biosignal Acquisition Platform based on Arduino and Android 21
Identification of benign breasts during mammogram screening 23
Color feature selection for unconstrained iris recognition 25
Automatic assessment of Leishmania infection indexes of in vitro macrophage cell cultures 27
Local interest detector based cancer cell mobility and morphology joint analysis 29
Vital Analysis: a framework for annotating physiological signals of first responders in action 31
An Automatic Graph-based Method for Retinal Blood Vessel Classification 33
Color image processing as a monitoring tool in gas combustion systems 35
The importance of color spaces in robotic vision 37
Application of global field power and SVM-RFE on affective computing 39
Classification of Gastroenterology images using Invariant Gabor Texture features 41
Exploring EMD for Lung Crackle Detection 43
Geometry Construction Recognition by the Use of Semantic Graphs 45
Compression of DNA microarrays using a mixture of finite-context models 47
Bone Contour Segmentation from US Images – A Comparative Study 49
Segmentation of the Lungs in PET Scans: A Watershed-based Approach 51
Study of cerebrovascular dynamics using functional MRI 53
A Total Variation based Denoising Algorithm for 3D Ultrasound 55
Estimating world coordinates in perspective vision systems for humanoid robots 57
Robotic Soccer: a real challenge for cooperative robotics 59
A platform for segmenting and characterizing polygonal networks on remotely sensed images 61
Unravelling the molecular determinants behind the tumour spectrum associated to TSG: CDH1 and HDGC as model 63
Characterization of the intronic portion of cadherin superfamily members 65
Lane Background Removal in Thin-Layer Chromatography Images Using Continuous Wavelet Transform 67
Motion Flow Trajectories: The Beginning of a Visual Perceptual Reasoning System in a Unified Data-Driven Approach 69
E-cadherin spatial characterization with radial distribution profile for mutation detection. 71
Assisted Teleoperation of a Quadrotor using Active Perception 73
Using Phoneme Acoustic Models for Fast Word Spotting 75
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word count variation in 1,092 human genomes</td>
<td>79</td>
</tr>
<tr>
<td>Evaluation of different incremental learning methods for video surveillance scenarios</td>
<td>81</td>
</tr>
<tr>
<td>On the compression of FASTQ quality-scores</td>
<td>83</td>
</tr>
<tr>
<td>A Two-stage Mass Segmentation Method for Breast Ultrasound Images</td>
<td>85</td>
</tr>
<tr>
<td>A PCA Approach to ECG-based Biometrics</td>
<td>87</td>
</tr>
<tr>
<td>Use of Artificial Neural Networks for Identifying Forces on a Bulldozer Ripper</td>
<td>89</td>
</tr>
<tr>
<td>Extracting BI-RADS features from mammography report</td>
<td>91</td>
</tr>
<tr>
<td>Breast Ultrasound Gland Segmentation</td>
<td>93</td>
</tr>
<tr>
<td>Extracting Health Information From Portuguese Tweets</td>
<td>95</td>
</tr>
<tr>
<td>Simulation of Ultrasound Guided Robotic Surgery with MORSE</td>
<td>97</td>
</tr>
<tr>
<td>Immersive Robot Teleoperation Using an Hybrid Virtual and Real Stereo Camera Attitude Control</td>
<td>99</td>
</tr>
<tr>
<td>Ultrasound quantification of Steatosis in patients with chronic infection of hepatitis C virus</td>
<td>101</td>
</tr>
<tr>
<td>Detection of Children Speech Disfluency</td>
<td>103</td>
</tr>
<tr>
<td>Effect of Image Resolution in DLT-Lines Camera Calibration Including Radial Distortion</td>
<td>105</td>
</tr>
<tr>
<td>3D Point Cloud Registration of the Femoral Bone, using the Point Cloud Library</td>
<td>107</td>
</tr>
<tr>
<td>Shots Detection of Sport Headlines using the Angular Orientation Partition Descriptor</td>
<td>109</td>
</tr>
<tr>
<td>3D reconstruction of bat trajectory from stereo vision</td>
<td>111</td>
</tr>
<tr>
<td>Actigraphy movement classification for sleep/wakefulness discrimination</td>
<td>113</td>
</tr>
<tr>
<td>An incremental linear model for the dynamics of the sinoatrial node</td>
<td>115</td>
</tr>
<tr>
<td>Towards the Evaluation of a Fuzzy Logic Model to Describe SBP and RR Relation: a Case Study Report</td>
<td>117</td>
</tr>
<tr>
<td>Predicting Malignancy from Mammography Findings and Surgical Biopsies</td>
<td>119</td>
</tr>
<tr>
<td>Automatic Model to Infer Hypertension in Children</td>
<td>121</td>
</tr>
<tr>
<td>Detection of Carotid Plaque Symptoms using Ultrasound Imaging</td>
<td>123</td>
</tr>
<tr>
<td>Graphical criteria of ROC analysis for endarterectomy purposes</td>
<td>125</td>
</tr>
<tr>
<td>Facial Dynamics for Identity and Expression Recognition</td>
<td>127</td>
</tr>
<tr>
<td>Detecting Cardiac Pathologies from Annotated Auscultations</td>
<td>129</td>
</tr>
<tr>
<td>Fast performance 3D object recognition: Dealing with rotationally symmetric objects</td>
<td>131</td>
</tr>
<tr>
<td>Comparative Study of Inverse and Forward Sensor Models in Occupancy Grid Mapping Using Sonars</td>
<td>133</td>
</tr>
<tr>
<td>An Eye2Eye Human Interaction Evaluation</td>
<td>135</td>
</tr>
<tr>
<td>Hand Pose Estimation using a Top-Down Bottom-Up based Approach</td>
<td>137</td>
</tr>
<tr>
<td>Computer-Based Classification of Cutaneous Melanocytic Lesions in Dermoscopic Images</td>
<td>139</td>
</tr>
<tr>
<td>A Signal Acquisition and Processing Device to Assist Human Heart Sound-based Diagnosis</td>
<td>141</td>
</tr>
<tr>
<td>Applications for people with special needs using Kinect</td>
<td>143</td>
</tr>
</tbody>
</table>
INVITED SPEAKER
Invited Talk

Indoor Navigation Infrastructure based on Augmented Reality Techniques

Carlos Gonzalez-Morcillo
Associate Professor of Computer Science at the University of Castilla-La Mancha (Spain).
Carlos.Gonzalez@uclm.es
cglez@inf-cr.uclm.es

Abstract:
Indoor navigation systems have represented a hot research topic for the research community with many different proposals to position people and keep track of their movements. In this lecture, Dr. Morcillo will take a step forward describing an object-oriented distributed architecture for highly scalable indoor navigation systems. The idea behind this architecture is to assist people with different needs while they stay in large spaces or buildings, such as public administration buildings, transport facilities, hospitals and so on, helping them reach their goals in such environments.

In this context, the use of Augmented Reality techniques boosts the concept of mobility thanks to the underlying architecture, integrating heterogeneous hardware devices and tracking methods. The resulting platform is a multi-layered scalable architecture based on autonomous agents. The most relevant work carried out by the Applied Artificial Intelligence Research Group at the University of Castilla-La Mancha will be also discussed.

Carlos Gonzalez-Morcillo is an Associate Professor of Computer Science at the University of Castilla-La Mancha (Spain). He received the BsC and PhD degrees in Computer Science from the University of Castilla-La Mancha in 2002 and 2007, respectively. His research interests include Distributed Rendering, Augmented Reality, MultiAgent Systems, and Intelligent Surveillance. Dr Morcillo has worked in the fields of Computer Vision and Data Mining at the Software Competence Center Hagenberg (Austria). He is Blender Foundation Certified Trainer and member of the Eurographics Society. He is also co-author of the Spanish book Fundamentals of 3D Image Synthesis, a practical approach with Blender. Further information can be found at http://www.esi.uclm.es/www/cglez.
PAPERS
Abstract

Breast Cancer Conservative Treatment (BCCT) is established as the most used procedure for breast cancer treatment. The use of this technique made essential the aesthetic outcome evaluation. Surgical outcome depends on several factors, leading to significantly heterogeneous results, due to the difficult to assess them. The limited reproducibility of subjective evaluation, led to the development of different techniques based on objective methodologies, such as Breast Cancer Conservative Treatment. The tool software. Recently, several studies proved that the introduction of simple three-dimension (3D) approaches could improve the performance of BCCT, maintaining the simplicity and concept behind the tool. The introduction of this new type of data, led to new challenges on the detection of important characteristics.

The objective of this work is to develop a simultaneous detection of breast contour and breast peak point, in images acquired using a depth-based low-cost system. Experimental results show the suitability of our depth-map based approach. The proposed algorithm has proven accurate and robust for a wide number of patients. Additionally, in comparison with previous research, the procedure for detecting prominent points was automated.

1 Introduction

Breast conservative therapeutic methodologies aim to obtain, local tumour control and survival rates equivalent to mastectomy, with better aesthetic results. However, the cosmetic outcome does not yet have an evaluation standard [2].

Until recently, the aesthetic evaluation was performed subjectively by one or more observers either by directly observing the patient or photographs, using scales which compares the treated with non-treated breasts. The most widespread scale is the Harvard scale, introduced by Jay Harris in 1979, which classifies the overall cosmetic results in four classes from excellent, good, fair to poor. Soon became clear that this kind of subjective evaluation had significant disadvantages. For example, exception is not guaranteed, reproducibility is difficult to attain and the level of agreement between observers is low or moderate.

Objective methods were introduced in an attempt to overcome the lack of objectivity and reproducibility. These methods compare the two breasts with simple measurements marked directly on the patients or on photographs of them. BCCT.core [1] is a recent computer-aided tool, which predicts the overall cosmetic result using features semi-automatically extracted from frontal photographs of patients. These features captures some of the factors which are considered to have an impact on the overall cosmetic results: breast asymmetry, skin colour changes caused by radiotherapy and surgical scar [1]. Although it is innovative and reproducible, this tool has important setbacks that are related to the complete automation of the software – which is fundamental for high reproducibility – and the ability to extract volumetric information to improve the overall cosmetic evaluation.

Several research groups have recently made efforts with 3D technologies [3, 4]. However, current 3D technologies are very costly and difficult to operate, thus requiring specialized staff and are not a feasible choice for daily clinical practice, therefore, its widespread use in the near future is not predictable.

Recently, we introduced Microsoft Kinect [5] as a promising low-cost and easy to use tool to evaluate the aesthetic result of BCCT, because it can not only simplify automation, but also provide volumetric data.

The goal of this paper is to model the mutual context of breast contour and breast peak (the area in the breast closer to the camera or further away from the chest wall, not necessarily the nipple) so that each can facilitate the recognition of the other.

2 Detection of Breast Contour and Breast Peak Points

Using context to aid visual recognition is recently receiving more and more attention. Context plays an important role in recognition in the human visual system, with many important visual recognition tasks critically relying on it.

When performed independently (breast contour and breast peak points detection), both tasks are non-trivial since many other parts of the image may be falsely detected. However, the two tasks can benefit greatly from serving as context for each other [6]. The algorithm was implemented with the following operations: 1) background segmentation; 2) breast peak candidates detection; 3) contour detection; 4) peak and breast contour decision. The simultaneously detection of the peak and breast contour will be addressed first by over-detecting peak candidates, followed by a contour detection near them.

Although the background should preferably be uniform for the acquisition process, sometimes it is cluttered. The presence of different objects at different depths, possibly at depths similar to the patient’s, dismisses the application of simple thresholding methods, such as Otsu’s (see Fig. 1(c)). To solve this problem we admit that the patient is in a somewhat central position in the image and it is likely the ‘object’ closer to the camera. A density image was defined by transforming the depth information on the XY plane to the XZ plane, where the value at position \((x, z)\) represents the histogram of the column \(x\), by counting along the \(Y\) direction (see Fig. 1(a)). Then for each \((x, z)\) position we computed the variance above and below. Each column \(x\) presents 3 different patterns: (1) background; (2) ‘object’ and background; (3) ‘object’. The XZ image is then replaced by the following rule: (1) cumulative value of the minimum of the variance from 1 to \(Nbins\); (2) average of the two variances; (3) cumulative value of the minimum of the variance from \(Nbins\) to 1 (see Fig. 1(b)). A global thresholding method of the original XY image corresponds to defining a horizontal line in the XZ image, discriminating background from foreground (see Fig. 1(a) and Fig. 1(c)). An adaptive thresholding method can be defined as a curve in the XZ image from left to right margins. This results in a threshold that varies from column to column in the original XY image. Since it is necessary for the curve in the XZ image to avoid the parts of the image with high values (high counts), the threshold curve was computed as the shortest path from left to right margin, where the cost of each pixel is its ‘intensity’ value (see Fig. 1(b) and Fig. 1(d)).

Figure 1: a) XZ plane depth information; b) XY variance plane; c) Otsu’s segmentation; d) Adaptive segmentation.

To model the breast peak point, a filter is used which evaluates the degree of divergence of the gradient vectors within its region of support from a pixel of interest (see Fig. 2). The peak point of the breast corresponds to the point in the breast where disparity attains the lowest value. The typically round or tear drop shape of a breast, leads to a distinctive pat-
tern in the gradient vector field where the gradient diverge in all directions (see Fig. 2(a)). Breast peak candidates were detected based on all local maximum positions, assessing the similarity between the template (see Fig. 2(b)) and the image using two different measures: cross-correlation \((f \ast g)[n] = \sum f^\ast[m]g[n + m]\), where \(f^\ast\) denotes the complex conjugate of \(f\); and circular correlation introduced by Nicholas Fisher in 1983.

![Breast gradient vector field](image)

**Figure 2:** (a) Breast gradient vector field (5-pixel spacing); (b) Template vector field.

Breast contour detection was performed using a shortest path approach. Intuitively, breast boundary manifests itself as a change in the grey-level values of the pixels, thus originating an edge in the resulting image. Therefore, interpreting the image as a graph with each pixel as a node and edges connecting adjacent pixels, the breast contour corresponds to a low-cost path through pixels, with the appropriate weight function. Since the breast contour is approximately circular and centered on the breast peak candidates, the computation is more naturally performed by adopting polar coordinates, with the origin of the coordinates in the peak candidate. Each column in the polar image corresponds to the gradient along each radial line in the original space, computed using a 3-point numerical differentiation: \(G_\theta(r) = \frac{f(r + h) - f(r-h)}{2h}\), where \(h = 1\) and \(r\) is the radius. Then, the gradient image is considered as a weighted graph with pixels as nodes and edges connecting neighbouring pixels. Each 4-neighbour pixel arc corresponds to a weight determined by the gradient value of the two incident pixels, expressed as an exponential law: \(f(g) = f_1 + (f_2 - f_1)\exp[\beta(255 - g) - 1]\), with \(f_1, f_2, \beta \in \mathbb{R}\) and \(g\) is the minimum of the gradient computed on the two incident pixels. For 8-neighbour pixels the weight was set to \(\sqrt{2}\) times that value. The parameters \(f_1, f_2, \text{ and } \beta\) were fixed at \(f_2 = 2, f_0 = 128, \beta = 0.0208\).

In this work, we are mainly interested in obtaining the localization of the breast contour, not so much in its complete delineation. Therefore, the angle \(\theta\) was varied only between \(\pi\) and \(2\pi\) (see Fig. 3(a)). The candidate contour was then the output of the shortest path algorithm in the polar image (see Fig. 3(b)). The shortest path was computed between the whole external margin and a single point (point of highest gradient) in the internal margin.

![Breast contour - ground truth](image)

**Figure 3:** (a) Breast contour - ground truth (solid red line), detected (dashed white line); (b) Polar image and detected contour (white line).

The quality or probability of the join model \((Q(CP) = \mu(DC) \cdot \rho_P)\) for the co-occurrence of breast peak and breast contour will be proportional to the individual qualities of the two parts, where \(\mu(DC)\) is the mean gradient alongside the contour and \(\rho_P\) is the cross-correlation of the centre candidate. Therefore, the final decision consists in selecting the pair (peak, contour) that maximizes the quality measure.

## 3 Results

The database now consists of 144 cases (depth images with \(640 \times 480\) px). Manual ground truth annotation was performed both to breast peaks position and breast contour definition. The breast peak points detection accuracy was measured using Euclidean metric distance (see Table 1).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Breast</th>
<th>Simultaneous detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>6.68 (3.60)</td>
<td>6.88 (3.90)</td>
</tr>
<tr>
<td>Right</td>
<td>13.61 (39.02)</td>
<td>10.32 (27.51)</td>
</tr>
</tbody>
</table>

![Table 2: Breast contour detection error (in pixels).](image)

First column (standalone) shows the detection error that would be obtained by grounding the decision on the maximization of the output of the convergence filter. The second column (simultaneous detection) depicted the performance for the proposed scheme. Miss detection means that breast peak point was detected outside from breast area. It is clear the advantage of the proposed method. Moreover the cross-correlation attains better results both in the mean error and in the miss detection.

Breast contour detection error was evaluated based on the Hausdorff and the average distances (Table 2). The Hausdorff distance is defined as \(h(A,B) = \max \| a - b \|\), where \(B\) represents the set of pixels of the ground truth and \(A\) the segmented breast contour; \(\| \cdot \|\) is the Euclidean distance. The motivation for using this metric is that it represents the worst case scenario.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Breast</th>
<th>Simultaneous detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>6.68 (3.60)</td>
<td>6.88 (3.90)</td>
</tr>
<tr>
<td>Right</td>
<td>13.61 (39.02)</td>
<td>10.32 (27.51)</td>
</tr>
</tbody>
</table>

## 4 Conclusions

In this paper we presented the simultaneous detection of prominent points on the breast using depth-map data acquired with a Microsoft Kinect sensor. Breast peak points were found based on gradient vector field information associated with the convergence filter theory. Breast contour was found as the solution to the shortest path problem is the graph theory framework, after conveniently modelling the image as a weighted graph. It was shown that depth-map images facilitate the automation of BCCT core, keeping this software low-cost and easy to perform. The results obtained also indicate an excellent performance and robustness for a wide variety of patients. Since the Kinect resolution is 1.3mm/pixel, the average error corresponds to 2.36mm. Future work will focus on the detection of the complete breast contour, including start and end points, conversion of prominent points to colour images, automatic nipples detection and extraction of volumetric information.

## 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 PCOMP - /01-0124-FEDER-022701 PhD grant with reference SFRH/BD/43772/2008 and project PTD/C/SAU-ENB/114951/2009.

## References

Music Symbols Segmentation through Recognition

Ana Rebelo\textsuperscript{1}
arebelo@inescporto.pt

Cuihong Wen\textsuperscript{2}
cuihongwen2006@gmail.com

Jaime S. Cardoso\textsuperscript{1}
jaime.cardoso@inescporto.pt

\textsuperscript{1}INESC TEC and Faculdade de Engenharia Universidade do Porto, Portugal
\textsuperscript{2}Hunan University Changsha, Hunan, China

Abstract

Optical music recognition (OMR) systems have been under intensive development for many years in order to create a robust process for printed and handwritten music scores. In this paper, a method to extract the music symbols from a music sheet without segmentation is presented. The aim is to execute simultaneously the segmentation and recognition of the objects avoiding the issues inherent to the segmentation phase. A Combined Neural Network (CNN) framework is also proposed to classify the music symbols.

1 Introduction

The musical symbols detection is a stage on an OMR system where operations to localize and to isolate musical objects are applied. In [4] a process to segment the objects based on a hierarchical decomposition of a music image and in contextual information and music writing rules was proposed. First the image was segmented in order to detect and isolate the primitive elements and then the symbols were classified. In this work, a new method is presented. The idea is to perform segmentation through recognition, that is, the method simultaneously segment and recognize the image. The principal advantage regarding to the previous method is in the elimination of the multiple heuristics used in the first case. As symbols are first detected and extracted from the image and, after that, classified, various parameters related to the size, shape and position of the objects are introduced. These parameters can constitute a severe problem in handwritten music scores, because the variability in writing style of each composer. Segmenting the music sheet using classification simplifies all the process and also overcomes the issues inherent in sequential detection of the objects, leading to less errors.

2 Detection Process

The recognition process consists first in the splitting by staffs the music sheet with staff lines previously removed [2], and then analyzing each of these segments. For each staff the connected component technique is applied. This technique has a threshold in order to join neighboring pixels from broken objects. The algorithm proceeds with a scanning of each connected component in order to recognize what is an object and what is not. This analysis is carried out using classifiers and it is a hierarchical process. On the first level of this hierarchy, the detected objects are split into note and symbol, and then, on the second level, the noise objects are divided into four types (see Figure 1): (1) connected symbol, (2) not symbol, (3) split symbol and (4) connected and split symbol. The objects that are classified as symbol can be one of the 19 possible symbols presented in Table 1.

Table 1: The set of the musical symbols considered in the symbol class.

<table>
<thead>
<tr>
<th>Accent</th>
<th>Dot</th>
<th>Bass Clef</th>
<th>Treble Clef</th>
<th>Flat</th>
<th>Natural</th>
<th>Note</th>
<th>Note Flag</th>
<th>Note Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>‹</td>
<td>⋆</td>
<td>‹</td>
<td>■</td>
<td>†</td>
<td>†</td>
<td>†</td>
<td>†</td>
<td>†</td>
</tr>
<tr>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td></td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
</tr>
<tr>
<td></td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
</tr>
</tbody>
</table>

On the third level of the hierarchical classification process, the algorithm scans again each connected component that is one of the four types of noise objects. These objects, presented in Table 2, can be classified in one of the 14 possible symbols to find. Note that in this step we are trying to split objects, usually notes connected to beams, so the waste class is necessary.

Table 2: The set of the musical symbols considered in the waste class.

<table>
<thead>
<tr>
<th>Accent</th>
<th>Dot</th>
<th>Bass Clef</th>
<th>Treble Clef</th>
<th>Flat</th>
<th>Natural</th>
<th>Note</th>
<th>Note Flag</th>
<th>Note Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>‹</td>
<td>⋆</td>
<td>‹</td>
<td>■</td>
<td>†</td>
<td>†</td>
<td>†</td>
<td>†</td>
<td>†</td>
</tr>
<tr>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
<td>Rest</td>
</tr>
<tr>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
</tr>
<tr>
<td></td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
<td>⋆</td>
</tr>
</tbody>
</table>

Figure 1: Examples of objects that are considered noise.

As mentioned in the beginning of this section, the analysis executed during the procedure to select the musical symbols relies on classifiers. We built four types of classifiers relating them to each possible situation that can appear in the connected components. The CNN_1 classifier \textsuperscript{1} is performed in order to divide the objects detected as noise or symbol. If the detected object is symbol then the CNN_3 is used, otherwise the CNN_2 is utilized. For connected symbol, split symbol and connected and split symbol an analyse to each bounding box of the object is carried out using the CNN_4. The construction of these classifiers will be explained later.

Connected symbol

Connected symbol class encompasses notes connected to beams, notes connected to accidentals and notes connected to accents. In the first situation accidentals and accents can also appear in the bounding box. We proposed a sliding window procedure supported by the CNN_4 to detect and extract the symbols. First, the analysis window with an height equal to the height of the bounding box is moved along the columns. The window width starts equal to \textit{staffspaceheight}\textsuperscript{2} and is scaled three times. The choice of using this value was obtained experimentally. Only the notes class is considered on this step. After this the search of the window is changed to go by rows. The aim is to look for accidentals and accents. Again the CNN_4 classifier is used to detect and extract the symbols. The procedure to establish the window size is the same of the previous step. Since an overlap exists between windows, there are repeated objects that need to be removed. Hence, a process to group symbols is executed. The symbols from the same class are compared with each other; if their positions are close enough, they are saved as one symbol. All the symbols detected are removed from the image. Now it is necessary to detect beams

\textsuperscript{1}The Combined Neural Network (CNN) will be explained on the next section.

\textsuperscript{2}\textit{staffspaceheight} represents the distance between two consecutive stafines. For more details please see [1].
which are music symbols linking two notes. So, for each image composed by two adjacent notes the algorithm looks for black pixels. It is worth restating that notes were already removed from the image.

**Split symbol**

Split symbol class encompasses broken objects. Usually they are notes separated from their stems or fragmented accidentals and clefs. The goal is to join black pixels near to the initial object. For that the window size increases until a certain limit and the CNN 4 recognizer is used to see when we are in presence of a music symbol. The augmentation of the window is first done in height, then in width and then in both. At the end the procedure to look for repeated symbols is again computed.

**Connected and split symbol**

Connected and split symbol class encompasses the two previous groups of symbols. For that reason, the techniques already described for each of the classes are applied here.

After the detection of all symbols a process to test some musical rules is executed. In here, the presence of accents only above notes and the position of accidentals before and at same height of notes is verified. If the symbols do not respect these rules then they are eliminated.

### 3 Combined Neural Network

To perform the various necessary classifications during the scanning procedure, we propose a music symbol recognizer based on a majority vote combination of three Multi-Layer Perception (MLP) classifiers named Combined Neural Networks (CNN). Two of the networks have the same architecture, but the initial random weights are different. The third network is fed with a different input and with a different number of neurons in the hidden layer. All these networks have a log-sigmoid activation function. In this way we expect to increase the overall performance of the classifier regarding to the usual way of only one MLP.

For two of the networks each image of a symbol was initially resized to 20 × 20 pixels and then converted to a vector of 400 binary values. For the third network each image of a symbol was initially resized to 60 × 20 pixels and then converted to a vector of 1200 binary values. Usually the images have an height larger than their width and the idea was thus to favor the height. In this manner, the problem in the classification of barlines, due to its similarity with dots after the resize, is minimized.

A database of training patterns was created according to the possible objects that algorithm could find in the scanning process. Also because of that, for each CNN we have a different database. Each one of these databases was randomly split into training and test sets, with 60% and 40% of the data, respectively. This division was repeated 10 times without restricting the distribution of the categories of symbols over the training and test sets. Only two constraints were imposed: at least one example of each category should be presented in the training set and the size of the noise class should be limited to avoid very unbalanced classes distributions. The best parametrization of each model was found using the training and validation sets, with the expected error estimated on the test set by a 4-cross validation scheme. The results for the different models can be seen in Table 3.

<table>
<thead>
<tr>
<th>Noise/Symbol</th>
<th>Noise Symbol</th>
<th>Noise Symbol</th>
<th>Noise Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>[91, 92]</td>
<td>[82, 84]</td>
<td>[68, 89]</td>
</tr>
<tr>
<td>CNN</td>
<td>[95, 96]</td>
<td>[90, 91]</td>
<td>[95, 96]</td>
</tr>
</tbody>
</table>

Table 3: 99% CI for the expected performance (in percentage) for the classification models.

If we compare with our initial results, with one MLP applied to the same datasets of music scores and also divided in the same way, we obtained an higher accuracy, as expected.

### 4 Experimental Testing and Conclusion

The data set adopted to test the proposed architectures for the music symbols extraction consists of both handwritten and synthetic scores. In total we have 9 scanned printed scores, 26 handwritten scores and 882 images generated from 18 synthetic scores (available from [3]). The metrics accuracy rate, average precision and recall were considered. They are given by

\[
\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{fn} + \text{tn}}, \quad \text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}, \quad \text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}
\]

The true positive rate (TPR), false positive rate (FPR), true negative rate (TNR) and false negative rate (FNR) were also considered:

\[
\text{TPR} = \frac{\text{tp}}{\text{tn} + \text{fn}}, \quad \text{FPR} = \frac{\text{fp}}{\text{tn} + \text{fp}}, \quad \text{TNR} = \frac{\text{tn}}{\text{tn} + \text{fp}}, \quad \text{FNR} = \frac{\text{fn}}{\text{tn} + \text{fn}}
\]

where \(tp\) are the true positives, \(tn\) are the true negatives, \(fn\) are the false negatives, \(fp\) are the false positives and \(tpc\) are the classes of the true positives. A false negative happens when the algorithm identifies a musical symbol as noise; and a false positive is when the algorithm identifies noise as a music symbol. These percentages are computed using the symbols position reference and the symbols position obtained by the segmentation algorithm.

<table>
<thead>
<tr>
<th>Handwritten Scores</th>
<th>Scanned Scores</th>
<th>Printed Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>Positive</td>
<td>91.7%</td>
<td>2.1%</td>
</tr>
<tr>
<td>Negative</td>
<td>99.0%</td>
<td>20.5%</td>
</tr>
</tbody>
</table>

Table 7: Confusion Matrix for the results from the Table 4.

Looking to the results, we can conclude that the algorithm to extract symbols through recognition detects more false negatives and less true positives than just using the algorithm for symbol extraction. This means that symbol extraction shows a reduction on the correct prediction of the symbols for the three datasets. This rationale is clearly depicted on the results shown in Table 6 where the recall substantially decreased over the recall results shown in Table 4. However, the first process has more missed symbols. Furthermore, comparing the Tables 4 and 6 the symbols extraction through recognition improves the performance in printed scores (82.2%). As future work, other experiments in the neural network could be addressed in order to improve the outcome: change the size of the third network each image of a symbol was initially resized to 60 × 20 pixels and then converted to a vector of 1200 binary values. Usually the images have an height larger than their width and the idea was thus to favor the height. In this manner, the problem in the classification of barlines, due to its similarity with dots after the resize, is minimized.

**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 projects FCOMP-01-0124-FEDER-022701, SFRH/BD/60359/2009 and PTDC/SAU-ENB/114951/2009.

### References


Denoising and Segmentation of the Second Heart Sound Using Matching Pursuit

F. Hedayioglu1  
fheday@dcc.fc.up.pt  
M. G. Jafari2  
mjafari@eecs.qmul.ac.uk  
S.S. Mattos3  
ssmattos@cardiol.br  
M. D. Plumbley2  
mark.plumbley@eecs.qmul.ac.uk  
M. T. Coimbra1  
mcoimbra@dcc.fc.up.pt  

1 Universidade do Porto  
Instituto de Telecomunicações,  
Porto, Portugal.  
2 Queen Mary University of London,  
Centre for Digital Music,  
London, United Kingdom.  
3 Unit of Fetal and Pediatric Cardiology,  
Royal Portuguese Hospital,  
Recife, Brazil

Abstract

We propose a denoising and segmentation technique for the second heart sound (S2). To denoise, Matching Pursuit (MP) was applied using a set of non-linear chirp signals as atoms. We show that the proposed method can be used to segment the phonocardiogram of the second heart sound into its two clinically meaningful components: the aortic (A2) and pulmonary (P2) components.

1 Introduction

Auscultation is one of the simplest, quickest and most cost effective techniques to listen to heart sounds. It is used to identify and diagnose a large number of heart conditions [1], making it an invaluable cardiac screening tool.

However, cardiac auscultation is a difficult skill to master, since heart sounds are highly complex in themselves. There are two main heart sounds S1 (first heart sound) and S2 (second heart sound). It has been recognised that S1 is comprised of up to four components produced during ventricular contraction. S2 is comprised of two main components [2]: the aortic component (A2), which is the sound produced by the closure of the aortic valve, and the pulmonary component (P2) (Figure 1), which is the sound produced by the closure of the pulmonary valve.

In addition, some cardiopathies and cardiac conditions, such as pulmonary hypertension can change the amplitude and delay of A2 and P2. Therefore the correct identification of these components can be an important indicator for screening and diagnosing some heart conditions and diseases.

Some work has been done to segment the A2 and P2 components of the second heart sound, among them Xu et al [3, 4] model the A2 and P2 components as nonlinear narrow-band chirp signals, with a fast decreasing instantaneous frequency over time. This method estimate the component with highest energy by using a visually designed 2D mask of the Wigner-Ville distribution to estimate its instantaneous frequency. Then reconstruct the component and subtract it from the signal, repeating this procedure until all components are estimated.

In [5] the authors also assumed statistical independency of A2 and P2 to perform blind source separation. They took advantage of the fact that in a normal auscultation procedure, the clinician sequentially listens to the heart sounds in four different locations. The periodicity of the heart sounds allowed the clinician to align the recordings of each location and simulate parallel recordings to be able to apply blind source separation and extract the A2-P2 interval.

In this paper we propose a technique to denoise the second heart sound (S2) by using matching pursuit (MP) using physiologically based time-frequency atoms. The second objective of this paper is to propose a method that can automatically detect A2 and P2 on the approximated S2 generated by the MP algorithm.

2 MATCHING PURSUIT

Matching Pursuit (MP) is an algorithm used to decompose a signal into a linear expansion of waveforms that are selected from a dictionary of time-frequency functions (called atoms) [6]. MP is a greedy method that iteratively adds a representation that is sparse in the dictionary D, i.e. only a few atoms participate in the approximations. The fact that MP is a sparse representation of the signal means that it can represent the underlying structures of the signal in a compact representation. Sparse representations are also effective when used for denoise [7].

The MP algorithm represents the signal s as a sum of weighted atoms φk from a dictionary D, i.e.:

\[ s(m) = \sum_{k=0}^{m} \alpha_k \phi_k \] (1)

Where \( \alpha_k \) are the weights of each \( \phi_k \) atom. The algorithm starts from an initial approximation \( s(0) = 0 \), and a residual \( r(0) = s \), it looks for the atom in the dictionary that has the highest dot product with the current residual. Once this atom is found, a scalar multiple of that atom is added so that \( s(k) = s(k-1) + \alpha_k \phi_k \), where \( \alpha_k = \langle r(k-1), \phi_k \rangle \) and \( r(k) = s - s(k) \).

The approximated signal \( s(k) \) is decomposed into a series of time-frequency atoms in decreasing energy order.

\[ s(m) = \sum_{k=0}^{m-1} \langle r(k), \phi_k \rangle \phi_k + r(m) \] (2)

where \( r(m) \) is the residual vector after \( s \) is approximated by \( m \) atoms, and \( \langle r(k), \phi_k \rangle > \) is the projection (weight) of \( \phi_k \) atom.

3 Proposed method

3.1 Signal approximation

After pre-filtering the heart sounds we apply the matching pursuit algorithm.

Since we are looking to match the data with its underlying physiological components. We used, as dictionary \( D \), a non-linear chirp signal model based on the modelling of Xu et al [3, 4] used to model A2 and P2. We assume that these atoms give a good approximation of the signal, since we are making assumptions about the underlying signals that compose S2. The atoms of the dictionary for the MP were made unit vectors, and their equations are as follows:

\[ g_w(s, u, f_1, f_2, t) = \frac{1}{\sqrt{t}} a(t) \sin\left(\frac{f_1 t}{10} + 2f_2 \sqrt{\frac{t}{10}} - 2f_2\right) \] (3)

\[ a(t) = (1 - e^{-t}) e^{-\frac{t}{\gamma}} \sin\left(\frac{\pi t}{60}\right) \quad 0 \leq t \leq 60ms \] (4)

Where \( s \) is the scale, \( u \) is the displacement (or shift), \( f_1 \) is the atom’s lowest frequency and \( f_2 \) is the atom’s highest frequency.

For each S2, we performed MP using two atoms as stop-condition, therefore, creating a signal that is an approximation of S2 (Figure 1).

3.2 Automatic detection of A2 and P2

The detection of A2 and P2 happens in the approximated signal \( s(k) \). Here we are trying to reproduce what the clinician is looking for: the two highest and more distinguished peaks in the second heart sound. Since we are assuming a healthy person, we can assume that A2 comes before P2. For each S2, we execute the Algorithm 1.
Algorithm 1 pseudo-code for detecting A2 and P2.

1. $\text{dif} = \frac{\delta S^2(t)}{\delta t}$
2. Look for changes in the signal of dif(peaks)
3. For all positive peaks:
   $\rho(i) = \int_a^b S^k(t) dt$, where $S^k(a) = 0 \land a < i$
   $S^k(b) = 0 \land b > i$
4. tmp = sort(p)
5. if tmp(0) is earlier than tmp(1)
   A2 = tmp(0)
   P2 = tmp(1)
6. else
   A2 = tmp(1)
   P2 = tmp(0)
7. endIf
8. return A2 and P2

4 Experimental results
4.1 Manual annotation
In normal patients, A2 is defined as being the earliest highest positive peak in S2, and P2 is the most prominent peak that follows A2.

We chose to make the annotations by identifying all the intervals in the recordings where A2 and P2 could be found. Then we compute the precise parameters of A2 and P2 (amplitude and position in time) by finding the positive peaks in the recording inside the identified intervals and measure its amplitude and time. Based on these amplitudes, we calculate the mean amplitude, the lowest observed amplitude and highest observed amplitude of each observed component (Equation 2). On each one of the four classical auscultation sites, the number of annotated S2 was 128. In total, 512 S2 were annotated. These annotations were made to be as short as possible.

Figure 2: Detected and annotated A2 and P2 on the recorded heart sound

5 Experimental results
The high correlation between the approximated S2 and the original ones shows that the approximated signal is a good approximation of the real one (Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
<th>Site D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>92.55%</td>
<td>95.00%</td>
<td>93.70%</td>
<td>88.71%</td>
</tr>
<tr>
<td>Min</td>
<td>81.61%</td>
<td>80.09%</td>
<td>83.62%</td>
<td>74.64%</td>
</tr>
<tr>
<td>Max</td>
<td>97.19%</td>
<td>98.33%</td>
<td>97.82%</td>
<td>96.66%</td>
</tr>
</tbody>
</table>

Table 1: The approximated S2 correlation with the original S2 in percentage, by auscultation spot.

The average squared error between the detected A2, P2 and the annotated A2, P2, per auscultation site (Table 2) shows that better estimations of P2 are made at the site B (second left intercostal space): this is, indeed, the best site to identify the pulmonary component [2].

Table 2: The approximated S2 mean squared error by auscultation spot.

<table>
<thead>
<tr>
<th></th>
<th>Site A</th>
<th>Site B</th>
<th>Site C</th>
<th>Site D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>0.0315</td>
<td>0.0079</td>
<td>0.0187</td>
<td>0.0428</td>
</tr>
<tr>
<td>P2</td>
<td>0.2181</td>
<td>0.0886</td>
<td>0.3038</td>
<td>0.1132</td>
</tr>
</tbody>
</table>

6 Conclusion
A new technique to denoise the second heart sound by using MP with non-linear chirp signals was proposed. We also demonstrated that the proposed method can successfully automatically segment A2 and P2.

This new technique has the potential to be used in a real clinical environment as a tool for improving the measurement of physiologically relevant components of heart sounds.

For future work, we plan to work on optimizations for processing time and lower the memory usage by using information of the time-frequency analysis of the recorded heart sound to lower the search space of the MP.

Despite being able to reasonably represent S2, our atoms seems to underestimate the value of A2 and overestimate P2. This may be due to the amplitude function in equation (4) not decaying quickly enough to reflect these changes.

References
A Relevance-based Linde-Buzo-Gray Approach for Supervised Feature Discretization

Artur J. Ferreira
arturf@isel.pt

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

Instituto de Telecomunicações, and
Instituto Superior de Engenharia de Lisboa
Lisboa, PORTUGAL

Abstract

In many learning problems, the use of feature discretization (FD) techniques attains adequate and compact representations of the data, using less memory, as compared to the original representation. Often, they lead to lower training time, while improving the classification accuracy. Many FD techniques, either unsupervised or supervised, can be found in the literature. In this paper, we extend and improve on previous work, with a new FD method based on the Linde-Buzo-Gray (LBG) algorithm, guided by a relevance criterion. The key advantage of our approach, as compared to previous ones, is that it can work in unsupervised, supervised, and semi-supervised problems, depending on the relevance criterion used to perform the feature sorting. Experimental results, on standard benchmark datasets with different types of data and learning problems, show the improvement of our method, as compared to other supervised FD approaches.

1 Introduction

A typical dataset with numeric features uses floating point or integer representations. In some cases, these features have noisy values or show minor fluctuations which are irrelevant or even harmful for the learning task. Moreover, some learning algorithms require a discrete representation of the data. In order to address these problems, feature discretization (FD) techniques have been considered extensively in the past [2, 4, 5, 7]. FD provides compact representations, with less memory usage, while at the same time often reducing the training time and improving the classification accuracy. The literature on FD contains many unsupervised and supervised techniques. To the best of our knowledge, there is no technique that can be adapted to work in unsupervised, supervised, or semi-supervised learning scenarios.

In this paper, we extend and improve on previous work [3] in which we have proposed an unsupervised FD technique based on the Linde-Buzo-Gray (LBG) algorithm [6]. The key novelty of this paper is that the proposed FD method uses a relevance criterion to guide the discretization process, thus being able to work in unsupervised, supervised, or semi-supervised learning scenarios.

In this paper, we extend and improve on our previous work [3, 4] in which we have proposed an unsupervised FD technique based on the Linde-Buzo-Gray (LBG) algorithm [6]. The key novelty of this paper is that the proposed FD method uses a relevance criterion to guide the discretization process, thus being able to work in unsupervised, supervised, and semi-supervised modes. Due to space constraints, this short paper focuses only on our experimental evaluation for the supervised case.

The remaining text is organized as follows. Section 2 reviews FD techniques. Section 3 presents our new FD method. Section 4 reports experimental results and Section 5 contains a few concluding remarks.

2 Feature Discretization

This Section briefly reviews FD techniques that have proven effective. We start by describing our previous approaches for unsupervised FD, in Subsection 2.1. We then review some supervised FD techniques in Subsection 2.2, against which we compare our proposed approach in Section 4.

2.1 Unsupervised Linde-Buzo-Gray

Recently, we have proposed an unsupervised scalar FD method [3, 6] based on the use of the Linde-Buzo-Gray (LBG) algorithm [6]. For a given number of discretization intervals, LBG discretizes the data under the criterion that the discrete representation has the minimum mean square error (MSE) relatively to the original representation. In our approach, named unsupervised LBG (U-LBG 1) and described as Algorithm 1, the LBG algorithm is applied individually to each feature and stopped when the MSE between the original and the discretized feature falls below a threshold $\Delta$ or when the maximum number of bits $q$ per feature is reached. Thus, a pair of input parameters $(\Delta, q)$ is necessary; choosing $\Delta$ equal to 5% of the range of each feature and $q \in \{2, 4, 8\}$ has proven adequate for many problems. We have also proposed a variant of U-LBG1, named U-LBG2, in which we use a fixed number of bits per feature $q$ (the only parameter of the algorithm, in this case). Both U-LBG1 and U-LBG2 rely on the idea that a discrete representation with low MSE is adequate for learning.

Algorithm 1 U-LBG1 - Unsupervised Linde-Buzo-Gray

Input: $X$, n × d matrix training set (n patterns, d features). Output: $\Delta$, $q$: maximum expected distortion and maximum number of bits/feature.

1. for $i = 1$ to $d$ do
2. for $b = 1$ to $q$ do
3. Apply LBG to the $i$-th feature to obtain a $b$-bit quantizer $Q_b(i)$;
4. Compute $MSE_i = \frac{1}{n} \sum_{j=1}^{n} (X_{ij} - Q_b(X_{ij}))^2$;
5. if $(MSE_i \leq \Delta$ or $b = q$) then
6. $Q(i) = Q_b(i)$; /* Store the quantizer. */
7. $\tilde{X}_i = Q(i)$; /* Quantize feature. */
8. break; /* Proceed to the next feature. */
9. end if
10. end for
11. end for

2.2 Supervised Techniques

In this Subsection we outline the main characteristics of three supervised FD techniques, that have proven effective in different problems. The information entropy minimization (IEM) [2] and its variant IEM variant (IEMV) [4] are successful supervised FD methods. They use information-theoretic criteria, namely the entropy minimization heuristic for discretization of a continuous value into multiple intervals, leading to small decision trees to classify the data. The class-attribute interdependence maximization (CAIM) [5] algorithm is based on a statistical approach, aiming to maximize the class-attribute interdependence and to minimize the number of discrete intervals.

3 Proposed Method: Relevance-based LBG

As in U-LBG1, our FD proposal, named relevance-based LBG (R-LBG) and described in Algorithm 2, uses the LBG algorithm, discretizing data with a variable number of bits per feature. The key difference is that now we use a relevance function, $\hat{rel}$ and a (non-negative) stopping factor, $\epsilon$. The relevance function, producing non-negative values, is applied after each discretization. Whenever there is an increase below $\epsilon$ on the relevance between two subsequent discrete versions (with $b$ and $b + 1$ bits), discretization is halted and $b$ bits are kept for that feature; otherwise, with a significant (larger than $\epsilon$) increase on the relevance, we discretize with one more bit, assessing the new relevance and repeating the relevance evaluation process. In summary, the key idea is to discretize a feature with an increasing number of bits, stopping only when there is no significant increase on the relevance of the feature.

The relevance assessment $r_b = \hat{rel}(Q_b(X_i); \ldots)$, of feature $i$ with $b$ bits, in line 5 of Algorithm 2 can refer to unsupervised, supervised, or semi-supervised learning. This depends on how the relevance function makes use (or not) of the available class labels. For instance, with $\hat{rel} = MSE$ (between original and discrete features) we have the unsupervised U-LBG1 approach. For the supervised case, we propose to compute relevance by the mutual information (MI) [1] between discretized features, $Q_b'(X_i)$, and the class label, $y$, by $r_b = MI(Q_b'(X_i); y)$ or by the
Algorithm 2 R-LBG - Relevance-based LBG

Input: $X$: $n \times d$ matrix training set $(n$ patterns, $d$ features), $y$: $n$-length vector with class labels (supervised mode), $q$: maximum number of bits per feature. 
@rel, $e$ $(>)$: non-negative relevance function and its stopping factor. 

Output: $X$: $n \times d$ matrix, discrete feature training set. 
$Q^i_1, \ldots, Q^i_q$: set of $q$ quantizers (one per feature).

1: for $i = 1$ to $d$ do
2: \hspace{1cm} $pRel = 0$; \hspace{1cm} $[^{[*]}$ Initial (and previous) relevance for feature $i$. $[^{*}]$]
3: for $j = 1$ to $q$ do
4: \hspace{2cm} Apply LBG to the $i$-th feature to obtain a $b$-bit quantizer $Q^i_j$; \hspace{1cm} $[^{[*]}$ Relevance-based LBG]
5: \hspace{2cm} Compute $r_{Q^i_j} = @rel(Q^i_j(X), \ldots)$, relevance of feature $i$ with $b$ bits; \hspace{1cm} $[^{[*]}$ Relevance computation]
6: \hspace{2cm} if ($r_{Q^i_j} - pRel < e$) then \hspace{1cm} $[^{[*]}$ Low increase. Discretize. Move to next feature. $[^{*}]$]
7: \hspace{2cm} $pRel = r_{Q^i_j}$; \hspace{1cm} $[^{[*]}$ High increase. Update relevance and add one bit. $[^{*}]$]
8: \hspace{2cm} end if
9: \hspace{1cm} end for
10: \hspace{1cm} end for
11: \hspace{1cm} $Q^i(\cdot) = Q^i_j(\cdot)$; \hspace{1cm} $\hat{X}_i = Q^i_j(X)$; \hspace{1cm} $[^{[*]}$ Store quantizer and discretize feature. $[^{*}]$]
12: end for

Well-known Fisher’s ratio (FIR). Figure 1 (top) shows the evolution of the MI (between features and the class label) on the R-LBG algorithm using $q \in \{1, \ldots, 10\}$ bit, on the discretization of features 1, 12, 16, and 18 of the Hepatitis dataset. In the bottom plot we compare MI using discretization with $q = 1$ and $q = 10$ bits. For the (real-valued) features 1, 16, and 18, the MI grows on the first few iterations using a small number of bits and then it levels off. In the case of the (categorical) feature 12, an increasing number of bits does not lead to a higher MI. Our method handles well.

4 Experimental Evaluation

We have carried out the evaluation of our method using publicly available datasets, with two-class and multi-class problems with different types of data and dimensionality (see Table 1). We evaluate the FD techniques (our R-LBG against the IEM, IEMV, and CAIM methods) regarding the average number of bits required to represent each instance and the classification accuracy of the 3 nearest neighbor (3NN) and linear support vector machine (SVM) classifiers (as implemented in the PRTools toolbox). The task is supervised classification, with the accuracy assessed by 10-fold cross-validation (CV). In each CV loop, we learn a quantizer in the training set and apply it to the test set.

Table 2 shows the average bits per instance whereas Table 3 presents the corresponding average test set error rate for both the 3NN and the linear SVM classifiers. For each classifier, the best result is shown in boldface and No FD corresponds to the original features. The CAIM algorithm takes a prohibitive time (several hours) on the higher-dimensional datasets (marked with symbol ‘*’).

5 Conclusions

In this paper, we have proposed a feature discretization technique based on the Linde-Buzo-Gray algorithm and guided by a relevance criterion, able to deal with unsupervised, supervised, and semi-supervised problems. The experimental results on classification accuracy show the improvement of our method as compared to other supervised FD approaches on multi-class high-dimensional datasets. On lower dimensional datasets, we obtain comparable results. As future work, we will apply the relevance criterion to perform feature selection on the discretized features.

References


Table 1: Datasets with $d$ features, $c$ classes, and $n$ instances.

Table 2: Average number of bits per instance, for a 10-fold CV, for the supervised FD methods on the datasets of Table 1.

Table 3: Average test set error rate (%) (3NN || linear SVM), for a 10-fold CV, for the supervised FD methods on the datasets of Table 1.

As compared to the other three FD techniques, usually the R-LBG method attains similar or lower number of discretization intervals. In many cases, it also leads to higher classification accuracy, specially on the multi-class high-dimensional datasets (#4 to #8). The R-LBG method shows promising results for multi-class high-dimensional data.
Supervised Feature Discretization by Mutual Information Maximization

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

Artur J. Ferreira
arturj@isel.pt

Abstract

Feature discretization (FD) techniques may be beneficial to many machine learning problems. FD leads to compact data representations, ignoring minor fluctuations that are irrelevant or even harmful for the learning task. Moreover, it is often the case that learning with discrete representations yields both lower training time and better accuracy. In this paper, we propose a supervised FD technique based on the maximization of the mutual information (MI) between each discrete feature and the class label. The discretization intervals are obtained incrementally using a recursive procedure. Experimental results on standard benchmarks, with binary and multi-class problems, show that our method usually achieves better accuracy than other well-known supervised FD approaches.

1 Introduction

In machine learning and data mining problems, a dataset is usually composed by categorical and numeric features. The categorical features are discrete by nature. In the case of numeric features, integer or floating point are common representations. For these features, the performance of machine learning and data mining tasks can often be improved by discretization/quantization. The discretization of numeric features leads to more compact representations, which ignore minor fluctuations on the data, leading to more robust classifiers. There are many unsupervised and supervised feature discretization (FD) techniques in the literature [3, 5, 7, 9, 11], the majority of which are supervised. Most supervised FD techniques use statistical or information-theoretic criteria to obtain the discretization intervals.

In this paper, we propose a new supervised FD technique based on a mutual information (MI) criterion. The key idea is to discretize each feature individually, in such a way that the boundaries of the non-uniform discretization intervals are computed to maximize the MI between the discretized features and the class label.

The remainder of this text is organized as follows. Section 2 reviews supervised FD techniques as well as their benefits. Section 3 presents our proposed MI-based method for FD. The experimental evaluation is reported in Section 4. Finally, Section 5 presents some conclusions.

2 Supervised Feature Discretization

This Section briefly reviews supervised FD techniques that have proven effective for many problems (Subsection 2.1); it also includes some brief remarks about learning with discrete features (Subsection 2.2).

2.1 Well-known Approaches

We now outline the main characteristics of four supervised FD techniques that have been proven effective in the past. This description is far from exhaustive, since FD is a topic with a long research history.

The information entropy minimization (IEM) [4] and its variant IEM variant (IEMV) [6] are successful supervised FD methods; these methods use information-theoretic criteria, namely the entropy minimization heuristic, for discretizing continuous value into multiple intervals, leading to small decision trees to classify the data. The class-attribute interdependence maximization (CAIM) [8] algorithm is based on a statistical approach, aiming to maximize the class-attribute interdependence and to minimize the number of discrete intervals. The class-attribute contingency coefficient (CACC) [10] adopts a static, global, and incremental approach; it shows good results regarding the number of discretization intervals and the training time of the classifiers.

2.2 Some Remarks About Discretization

In data mining and knowledge discovery, the use of discretization algorithms has played an important role [11]. These methods yield concise representation of continuous attributes, allowing the users to understand the data more easily and also making learning more accurate and faster. Besides the benefits for machine learning techniques, many authors have found that discretization can also help humans to better understand their data. Moreover, this happens with quite different types of data, such as dense microarray data or sparse data.

The possible drawback of discretizing data, as compared to the use of the original features, is mainly the time and memory usage taken by the discretization process. A detailed description of many FD methods, their resemblance and their key differences can be found in [3, 5, 7, 9] and the many references therein.

3 Mutual Information Discretization

In this Section, we present our proposed supervised FD method named mutual information discretization (MID). The key idea of the MID method is to discretize each numerical feature individually, trying to maximize the MI of the discretized feature with the class label. The method uses an incremental and recursive approach, by successively breaking each feature into intervals. Figure 1 depicts the idea of our approach using $q = 3$ bits yielding a 8-interval non-uniform quantizer. At level 1, we compute the optimal cut point which leads to the maximum MI between the binarized feature and the class label. Then, at level 2, for each partition, we compute a cut point under the same criterion. We repeat the process at level 3 (with 4 cut-points). The resulting non-uniform quantizer has 7 cut-points (8 intervals). The key motivations for this FD proposal are as follows:

- the MI between a feature and a class label is related to its accuracy for classification purposes; MI has also been found adequate for feature selection (FS) purposes in many methods (see [1] for a review of FS methods based on MI);

- at the discretization stage, we can search for discretization boundaries such that the resulting discrete feature has the highest MI as possible with the class label. Thus, by maximizing the MI at each partition and each cut-point, we are aiming at leveraging the classification accuracy of the discrete feature.

![Figure 1: Illustration of the incremental and recursive partition algorithm for feature discretization, using $q = 3$ bit leading to a 8-interval quantizer.](image-url)
In summary, given a training set with \( n \) instances and \( d \) features, \( X \), and a maximum number of bits per feature \( q \), the MID method proceeds as follows for each feature \( X_i \), \( i \in \{1, \ldots, d\} \): (a) categorical features are kept unchanged \( X_i = X_i \); (b) numerical features are recursively quantized using the MID method described above (Figure 1), yielding the quantizer \( Q(\cdot) \) and the discretized feature \( X_i = Q(X_i) \). The decision to treat a given feature as categorical is done by computing its arity; for instance, if the arity is less than 5, it is treated as categorical, otherwise, as numerical.

4 Experimental Evaluation

This Section reports the experimental evaluation of our method. This evaluation was carried out on standard benchmark datasets, with different types of problems and data, as described in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>( d )</th>
<th>( c )</th>
<th>( n )</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wine</td>
<td>13</td>
<td>3</td>
<td>178</td>
<td>Classify wine cultivar</td>
</tr>
<tr>
<td>2. Hepatitis</td>
<td>19</td>
<td>2</td>
<td>155</td>
<td>Hepatitis cancer detection</td>
</tr>
<tr>
<td>3. Ionosphere</td>
<td>34</td>
<td>2</td>
<td>351</td>
<td>Ionosphere signal return</td>
</tr>
<tr>
<td>4. Colon</td>
<td>2000</td>
<td>2</td>
<td>62</td>
<td>Colon cancer detection</td>
</tr>
<tr>
<td>5. AK10P</td>
<td>2400</td>
<td>10</td>
<td>130</td>
<td>Face classification</td>
</tr>
<tr>
<td>6. Brain-Tumor I</td>
<td>5920</td>
<td>5</td>
<td>90</td>
<td>Brain tumor classification</td>
</tr>
<tr>
<td>7. Prostate-Tumor</td>
<td>10509</td>
<td>2</td>
<td>102</td>
<td>Prostate cancer detection</td>
</tr>
</tbody>
</table>

Table 1: Datasets used in the experiments; \( d \), \( c \), and \( n \) are the number of features, classes, and instances, respectively.

Figure 2 (top) shows how the MI between features and class labels evolves as discretization is carried out with \( q \in \{1, 2, 3, 4\} \) bits for a random training partition of the Wine dataset. The training partition has class entropy \( H(C) = \log_2(3) = 1.585 \) bit/symbol; this will be the maximum value for the MI. Figure 2 (bottom) plots the evolution of MI for features 1, 7, and 8. We see an increase in the first few bits and then the values of MI level-off. The average MI values for all \( d = 13 \) features are 0.4002, 0.5793, 0.6386, and 0.6386, for \( q \in \{1, 2, 3, 4\} \) bits, respectively.

Our MID method is compared against three supervised FD methods, described in Subsection 2.1, namely IEM, IEMV, and CAIM. The classification accuracy of the 3 nearest neighbor (3NN) and the linear SVM classifiers, by the PRTools toolbox\(^1\), for a supervised classification task with a 10-fold cross-validation (CV) is reported. In each CV fold, we learn a quantizer in the training set and apply it to the test set.

We use the original features (no FD) and the features discretized by the FD methods, using \( q = 2 \) bit (Table 2) and \( q = 3 \) bit (Table 3). The best result is in bold face. On the higher-dimensional datasets the CAIM algorithm takes a prohibitive time (symbol '*'). In the majority of the experimental results reported in both Tables 2 and 3, the MID method attains the lowest error. The results in Table 2 are slightly better than those in Table 3, which suggests that, when using a fixed number of bits per feature, \( q = 2 \) is adequate for different types of data. Contrary to the CAIM and CACC methods, our approach scales well for high-dimensional data, regarding the running time. The IEM and IEMV methods also scale well for high-dimensional data, but MID usually leads to features with lower generalization error.

5 Conclusions

In this paper, we have proposed a supervised FD technique based on mutual information maximization between the discretized features and the class label. This technique is based on an incremental and recursive approach that finds the optimal cut points in the mutual information sense. 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, with different types of data. Our method also scales well for high-dimensional and multi-class problems.

As future work, we will modify our method in order to both allocate a variable number of bits per feature and to perform feature selection.

References


---

Table 2: Test set error rate (%) for 3NN || linear SVM classifiers, using \( q = 2 \) bit and arity \( a = 5 \), on a 10-fold CV, for the datasets in Table 1.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25.3</td>
<td>4.5</td>
<td>2.2</td>
<td>1.1</td>
<td>4.5</td>
</tr>
<tr>
<td>23.2</td>
<td>2.2</td>
<td>19.4</td>
<td>16.1</td>
<td>20.6</td>
</tr>
<tr>
<td>13.4</td>
<td>12.0</td>
<td>11.7</td>
<td>10.3</td>
<td>13.7</td>
</tr>
<tr>
<td>17.7</td>
<td>17.7</td>
<td>21.0</td>
<td>17.7</td>
<td>16.1</td>
</tr>
<tr>
<td>47.7</td>
<td>0.8</td>
<td>60.0</td>
<td>73.1</td>
<td>43.1</td>
</tr>
<tr>
<td>15.6</td>
<td>13.3</td>
<td>23.3</td>
<td>16.7</td>
<td>8.9</td>
</tr>
<tr>
<td>17.6</td>
<td>19.8</td>
<td>27.5</td>
<td>30.4</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Table 3: Test set error rate (%) for 3NN || linear SVM classifiers, using \( q = 3 \) bit and arity \( a = 5 \), on a 10-fold CV, for the datasets in Table 1.
Mobile Framework for Biosignal Acquisition and Remote Processing

Michel Cânovas\textsuperscript{1}  
michel.canovas@lx.it.pt  
Hugo Silva\textsuperscript{1}  
hsilva@lx.it.pt  
André Lourenço\textsuperscript{2}  
alourenco@deetc.isel.ipl.pt  
Filipe Canento\textsuperscript{1}  
fcanento@lx.it.pt  
Ana Fred\textsuperscript{1}  
alfred@lx.it.pt

\textsuperscript{1}Instituto de Telecomunicações  
Instituto Superior Técnico  
Avenida Rovisco Pais, 1, 1049-001 Lisboa, Portugal

\textsuperscript{2}Instituto Superior de Engenharia de Lisboa  
Rua Conselheiro Emídio Navarro, 1, 1959-007 Lisboa, Portugal

Abstract

A new age of biosignal processing and pattern recognition applications can emerge using the capabilities of standard mobile devices and smartphones. Their capabilities in terms of computational power, built-in sensors, and ability to connect to other peripherals through wireless technologies allows promising real-world applications. Despite all this, some advanced level of skill is needed in terms of the development of mobile applications, to actually benefit from these features. In this paper we present a framework for rapid prototyping, aimed at bridging this gap in the mobile development ecosystem. Our generic framework enables real-time data acquisition, processing, communication and visualization. An application used for ECG signals monitoring, entitled CardioWatch, is presented as proof of concept.

1 Introduction

Nowadays, smartphones play a fundamental role, enabling ubiquitous access to unprecedented volumes of near real-time information in an always-on/always-connected approach [9]. This ability, combined with their set of built-in sensors and the strengths of distributed computing, makes smartphones a pervasive tool for everyday tasks [11].

Mobile application development can be a thorough and time consuming process for researchers, which often falls out of the scope of their core interests. Furthermore, in tasks involving signal processing, once algorithms are coded, any changes require thorough revision of the application. Often, changes or hypothesis validation are evaluated in an offline manner using a scientific computing language, and then reflected on the mobile application. Thus, the main motivation behind our work is to create a tool which addresses real-time biomedical signal acquisition, processing, communication and display, taking advantage of mobile and web technologies. Our ultimate purpose is to give researchers, and health-related professionals a tool to easily develop mobile signal processing and communication applications tailored to their needs. These can be used in biomedical engineering and mobile monitoring of health, targeting the improvement of quality of care and independent living [5, 6, 9].

This platform seeks to enable the unobtrusive collection of large corpuses in real-world scenarios, which open new challenges in several pattern recognition domains.

2 System Architecture

The architecture is arranged in layers, decoupling the presentation from the processing counterpart. It follows a Model-View-Controller (MVC) architecture, as proposed by Silva et al. [7], namely, the UI is designed upon a web interface which is then bound to the native part of the application. Moreover, the definition of the application is made by assembling a network of Functional Blocks, that runs on top of a Workflow Manager (WFM). This network defines the behavior of the application, in such a way that each block performs a specialized task, ultimately resulting on a real-time information handling.

The WFM is the core of the application. It deals with the instantiation and control of the previously designed system, described according to a custom specification format proposed by our work, defined in JavaScript Object Notation (JSON). This structure uses a specification, which we called Data Processing Language (DPL), that is loaded and automatically instantiated at runtime by the application.

The UI is, as in most applications, one of the most important parts of the platform. It needs to be able not only to accurately display the data to the user, but also to get his inputs, all of this in intuitive and aesthetical ways. Another concern taken into account in the design of our framework was the unification of the layout of the platform, between the mobile and other clients (e.g. desktops). This led to the use of the Android OS WebView component, that enabled the interaction and display of HTML content, which also allows the use of exactly the same UI both on the mobile phone, and on a standard PC interface.

3 CardioWatch

To demonstrate the platform behavior, a demo application was built, entitled “CardioWatch”. The goal of this application is to process Electrocardiography (ECG) data measured by an external biosignal acquisition unit in a remote server, and show the heart rate of the subject. The acquisition unit is connected to the mobile device via Bluetooth. On the other end, there is a remote server that performs the signal processing, which is accessed over the web (see Figure 1).

Inside the Android-powered device this data flow is realized through a sequence of four blocks: a Bioflux, a FlotBlock and a HTMLField, which are, respectively, the virtual representation of: the physical acquisition device; the implemented socket connection to the server; the ECG line chart; and the heart rate (HR) numeric display. The DPL that supports this data flow it is described as follows:

![Figure 1: Architecture of our demo application, where biosignals are acquired from an external hardware device, displayed in a mobile device through a rich user interface, and sent to a remote server for efficient processing; the results of the processing stage are returned back to the phone for presentation to the user.](image-url)
Here we can see how the connections between the input (in) and output (out) channels of the blocks are performed, and also how the configuration parameters (fields under config) vary amongst the blocks. Each block extends the abstract Block class, although they perform distinct tasks.

For instance, the BioPlux block implements methods for controlling a bioPLUX [4] acquisition unit via Bluetooth, which is the device that gathers the ECG signal; the Websocket block is a block that implements a full-duplex single socket connection (W3C websocket [10]) over which messages can be sent/received to/from the remote server.

Listening for the server-returned information are the FlotBlock and the HTMLField blocks, which display the real time signal and HR. FlotBlock owes its name to the Javascript library used for plotting in JQuery, Flot, that produces graphical plots of arbitrary datasets on-the-fly on the client-side.

On the center of Figure 1 (in the tablet display) we show the main screen of the demo application. It has a very simple interface: the two central buttons start and stop the system; to change its configuration, the user must press the top left button featuring a gear-wheel.

The application connects automatically to the server and as soon as the start button is pressed, the Android device connects to the bioPLUX acquisition device and starts to receive data in real-time. The bottom panel displays the processed data.

### 4 Remote Pattern Recognition Processing

The signal is not acquired according conventional practices, where the electrodes are placed on the thorax, instead, the procedure follows the approach presented in [2], in which a 1-lead setup for ECG signal acquisition at the fingers using Ag/AgCl electrodes without gel is proposed.

This signal is acquired, as already mentioned, with a bioPLUX device, at a sampling rate of 1000 Hz. At the server side we used the sWSw framework [8], which provides a simple way to interface Python with external applications via WebSocket connection.

The raw data is received by the server and submitted to a 300 order band pass Finite Impulse Response (FIR) filter with a Hamming window, and cutoff frequencies of 5Hz to 20Hz, to enhance the ECG signal quality and increase the signal-to-noise ratio. These specifications, take into account the ECG information bandwidth, and empirical considerations extracted from the heuristic analysis of the signal. Concerning real-time applications, the signal is acquired and filtered in frames, and thus, a method to process each frame and merge them into a meaningful signal, is required. The way we chose to solve this problem was to use the overlap-add method [3].

This filtered signal is then downsampled by a factor of 10 before returning to the Android device, producing an amount of data more suited to the display capabilities of the mobile device and reducing the payload to be transferred.

For the QRS complex detection we used the modified real-time version of the Engelse and Zeelenberg (1979) algorithm, proposed in [1], which provides means for the calculation of the HR information to be returned for display to the user, based on the latest 10 QRS complexes.

This framework provides a rapid prototyping tested for complex algorithms in a mobile environment, that otherwise would require more time and effort to implement and update using more traditional programming approaches. This feature is especially useful when a whole set of algorithms are already available in a Matlab or Python toolbox, as is our case with the BioSPPy toolbox.

### 5 Conclusion and Future Work

This architecture has undoubted benefits, namely the ones arising from its versatility, due to the fact that the user can define his application behavior by combining the different blocks, and reuse the layout of a web UI.

Profiling shows that the system components running in the native part of Android OS have sub-millisecond performance. In fact, the bottleneck of the application is the Websocket block. For a sampling rate of 1000Hz, tests show that every data message exchanged with the server takes, in average, 110 ms since its creation until its retrieval, spending 108 ms in the roundtrip to the server alone (34 of those spent in the actual remote processing).

Although this mobile framework is part of a more complete system, it is able to run as a stand-alone application, provided that the DPL with the network of Functional Blocks only comprehends local processing blocks. The chosen architecture has proven to be holding the flexibility necessary for the intended purpose, with a minimal latency overhead.

The example application evaluated serves not only as a proof of concept, but also as a way to demonstrate the potential of our approach. The integration of the Websocket block to communicate with a remote server, provides a testbed for functionalities that can be supported by server-side algorithms. Future work will include the optimization of the Websocket block, and incorporation of server-side functionalities in new Functional Blocks to enable standalone operation and/or better use of the bandwidth of the communication.

### 6 Acknowledgments

This work was partially funded by the IT - Instituto de Telecomunicações under the grant "Android Biometrics Platform", and by the Fundação para a Ciência e Tecnologia (FCT) under grants SFRH/BD/65248/2009 and SFRH /PROTEC/49512/2009 whose support the authors gratefully acknowledge.

### References


Abstract

This paper presents a non freehand Ultrasound acquisition system for human femur reconstruction where the probe displacement and position are accurately controlled by an anthropomorphic arm robot. The aim of this system is to acquire a sequence of 2D parallel cross-sections evenly spaced along the displacement direction in order to perform an accurate 3D reconstruction of the femur to be used in the scope of computer-assisted orthopedic surgery. Here, the system is described as well as the image processing and calibration procedures implemented. Results with real data are presented to illustrate the operation of the system. The work presented here, was developed in the scope of the HIPROB project. This project consists on develop a robot for orthopedic surgery, solving some problems caused by current systems, that use fiducial markers in order to find the bone in space. With HIPROB is expected locate the bone using information from the ultrasound (US) images acquired in intra-operative and CT images acquired on pre-operative scenario.

1 Introduction

Robotic and computer-assisted technology are increasingly being developed for use in surgery to aid surgeons in providing more precision and accuracy, especially during procedures requiring fine movements. Moreover, the reproducibility is another important aspect to be considered with the goal to get better results. More than enough reasons for the growing number of robotic systems for medical applications in recent years, as example the DaVinci system used in laparotomy or [MAKOplasty] robotic system to orthopedic surgery. The idea of developing a robot system for positioning and movement an ultrasound (US) probe, emerges the need to obtain images spatially located and referenced to allow an more accurate volume reconstruction. Reconstruct a 3D volume with collected 2D images provides an excellent tool for intra-operative planning, but it is extremely difficult to achieve by freehand scanning. A robot-assisted ultrasound system would provide significant benefit by enabling more structured and optimized 3D ultrasound datasets, providing consistent reliable guidance to interventions, and automate the process. This system can be used always wants to do a scan with the US probe, but was thought to be used in image acquisition of femur bones and assist surgeons in orthopedic surgery to localize the bone in the space, through the registration between the US data and the CT information, obtained in pre-operative scenario. With this image-guidance system is possible to reduce or eliminating the use of fiducial markers, decreasing recovery time of the patient and consequences after surgery. This paper is organized as follows. In section 2, the concept of automatic acquisition of US images are presented. The next section describes the US image processing steps. Section 3 presents the experimental results. Finally, conclusions and future work are presented.

2 Robotized Ultrasound Image Acquisition

The robotic system used, consists of an Eurobotec IR52C robot manipulator, an ALOKA prosound 2 echograph with a 5 MHz probe and a computer with a standard video card for image acquisition. The probe is placed in the end effector of the robot, responsible for positioning the probe in contact with the femur. The images were acquired and the femur scanned with the best possible coupling at a constant speed. In the preparation of each experiment is necessary to define the pose of the robot that guarantees the best coupling between the probe and the femur, without hurting the patient. Initial and final points of acquisition must be defined, to perform trajectory planning along the femur. Whenever possible, trying to make linear trajectories. The estimation of the han-eye and the robot-world transformations, to know the positioning of the probe at every moment, is an important point of this work. These estimations allow to perform a precise three-dimensional reconstruction of the bone. The calibration matrix is obtained through the \( T_{cal} \) equation, according to the reference frames of figure 1.

\[
T_{cal} = T_0 \times T_1 \times T_2 \times T_3 \times P(u, v)
\]

(1)

where,

\[
P(u, v) = \begin{bmatrix}
S_x(u - u_0) \\
S_y(v - v_0) \\
0 \\
1
\end{bmatrix}
\]

\(P(u, v)\), represents the image frames, \(S_x\) and \(S_y\) are the scale factors for the \((u, v)\) pixel coordinates, obtained by a CIRS Ultrasound Calibration Phantom. This Phantom is a cube with a small egg and a large egg. There are two scanning surfaces and the targets are centered within the background material. Knowing the dimensions of each egg, it is possible to obtain the relationship \(\text{pixels/mm}\) for each level of depth of the US equipment and calculate the scale factors. \(u_0\) and \(v_0\) are the horizontal and vertical offsets of the US images.

3 Ultrasound Image Processing

Since the bone is a rigid anatomical structure, ultrasound signals are reflected, and the image only captures the top layer of the bone. Processing this type of images is a challenging task, since images are very noisy and blurred. The image quality decreases severely when approaching the knee, because there is less muscle mass. The bone gets closer to the probe, so the US reflections are more intense. Other issue is the difficulty of coupling the probe to the knee, reducing the image quality. Image Denoising was used to remove the noise that degrades the images, for example Speckle, the most common noise in US images. The objective is to smooth homogeneous areas while preserving the contours without distorting the images. The algorithm used is based on the maximum a posteriori (MAP) criterion with a total variation (TV) edge-preserving Gibbs prior. The method is formulated as an optimization task that is solved by the Sylvester equation, developed by Sanches et al. To compensate possible probe displacement errors and femur movements, Normalized Cross-Correlation was used to align consecutive images, through equation 4.
where, \( f \) is the image, \( \bar{I} \) is the mean of the feature and \( \overline{I}_{uv} \) is the mean of \( f(x,y) \) in the region under the feature. To extracting the bone in all images, several techniques are studied and implemented as the energetic methods based in Active Contour Models and probabilistic methods. In [4], previous version of this work, an edge-based model described as Geometric Active Contour [3], was used to segment the bone in the images, but this method failing to detect the exterior and interior boundaries when the initial contour is far from the desired object boundary. To overcome this problem a region-based active contour model proposed by Zhang et al. in [5], was used to segment the femur. The authors implemented an method that share the advantages of the well known region-based Chan-Vese method [1] and Geometric Active Contour models. The method utilizes the statistical information inside and outside the contour to construct a region-based signed pressure force (SPF) function in order to define the contour, described in equation (3). The Chan-Vese model is formulated by minimizing the following energy functional:

\[
E^{CV} = \lambda_1 \int_{inside(C)} |I(u,v) - c_1|^2 du dv + \lambda_2 \int_{outside(C)} |I(u,v) - c_2|^2 du dv
\]

where \( c_1 \) and \( c_2 \) are two constants which are the average intensities inside and outside the contour, respectively, obtained from minimization of equation (3).

\[
c_1 = \frac{\int_I I(u,v) \cdot H(\phi) du dv}{\int_I H(\phi) du dv}
\]

\[
c_2 = \frac{\int_I I(u,v) \cdot (1 - H(\phi)) du dv}{\int_I (1 - H(\phi)) du dv}
\]

The signed pressure force modulates the signs of the pressures forces inside and outside the region of interest so that the contour shrinks when outside the object, or expands when inside the object.

\[
spf(I(u,v)) = \frac{I(u,v) - c_1 + c_2}{\max(|I(u,v) - (c_1 + c_2)|)}
\]

4 Experimental Results

Many experiments were performed with this system, first with a cow femur bone in a water tank and after in human legs. For example the figure 2b represents an acquisition detail of an human leg. In this experiment performed, a 18[cm] girl leg was scanned, on the central part of the leg from the hip to the knee. 297 images 720 × 576 pixels were acquired, meaning that the spacing among images is 0.6061[mm]. Figure 2a, represents an US image of the central part of femur and b) the result after applied the Denoising algorithm. The segmentation method described in section 3 was applied to all images. The initialization method is defined with a curve around the bone, identified by the user in the first image. After initialization, the algorithm automatically evolved to maximize the cost function. Figure 3 represents the result of segmentation of an US image. With all segmented images the 3D volume is performed, with techniques of surface rendering. Figure 4 shows the 3D reconstruction surface of the cow femur bone and the reconstruction of central part of the femur, where the images were acquired.

5 Conclusions

This paper presents a system for 3D reconstruction of the human femur from parallel and evenly spaced US cross sections acquired with a robot arm. Using this procedure is guaranteed the same pose of the probe and maintain the 3D orientation of each image slice, when moving the robot along the femur. The location of each pixel of the image is obtained through a calibration process, that calculates the transformation matrices through direct kinematics of the robot, taking into account the probe attached to the robot hand rigidly. A method to Denoise the US images, using a Bayesian algorithm based on a multiplicative model for the noise, is a good solution to pre-process the images and facilitate the bone segmentation process. The region-based method used to segment the bone presents good results, besides having a lower computational cost than the previously used. As future work is expected to improve the processing times to real time applications and perform the registration between US images and CT images of femur bones.

References

Multimodal biometric recognition under unconstrained settings

João C. Monteiro  
joao.carlos.monteiro@fe.up.pt  
Jaime S. Cardoso  
jaime.cardoso@inescporto.pt

Abstract

Unimodal biometric systems, based on a single biometric trait such as fingerprints for example, suffer from a wide variety of problems such as noisy data, intra-class variations, non-universality, spoof attacks or unacceptable error rates. Such limitations can be overcome through the development of multimodal biometric systems that integrate data acquired from multiple sources of information.

In the proposed work we aim to investigate and develop a multimodal biometric recognition system based on fingerprints, iris and voice, to overcome the problems presented by recognition algorithms based on each of these traits alone. With such system we aim to obtain recognition ratios that surpass the existing state-of-the-art multimodal biometric systems. Special focus will be given to the robustness of the developed algorithms exposed to inputs acquired under non-ideal conditions.

1 Introduction

Biometric systems make use of the physiological and behavioral traits of individuals, such as fingerprints, palmprints, iris patterns, voice, face, for recognition purposes [6]. The use a single trait for recognition (unimodal biometric systems) is often affected by practical problems such as susceptibility of sensors to noise, proneness to interclass similarities within large population groups, non-universality of single traits and vulnerability to spoofing [10, 11]. Multimodal biometric systems overcome some of these problems by consolidating the evidence obtained from different sources [5]. These sources may be multiple sensors for the same biometric, multiple instances of the same biometric, multiple snapshots of the same biometric or multiple representations and matching algorithms for the same biometric [7]. Given the number of possible biometric traits and the amount of combinations and ways of performing information fusion, there is still a lot of work to develop in the area, namely in the adaptation of unimodal algorithms to less constrained acquisition environments and the quality assessment of acquired data.

After the unimodal algorithms are stable the fusion algorithms and integration in a multimodal system shall work to improve the recognition rates as well as identifying how “good” a specific modality is for the recognition process. This quality assessment process is important in multimodal biometric systems as some of the systems might be highly corrupted by noise and, thus, negatively affect the global recognition performance.

The main goal of the present work will be the development and evaluation of pattern recognition algorithms for application in several unimodal biometric systems using input signals acquired under unconstrained settings. Using quality assessment and information fusion algorithms a multimodal system with improved recognition performance will be developed.

2 State-of-the-art

Brunelli, in its 1995 pioneer work [2], proposed the use of multiple sensors to capture different biometric traits and develop a multimodal recognition system. This system, compared with single trait unimodal biometric systems, would allow better performance under noisy acquisition conditions (multiple modalities generate more data for matching), as well as lower intra-class similarity and less susceptibility to adversarial attacks (a random set of traits can be asked to the user) [9, 11].

With multiple unimodal systems already developed and commercialized, works in multimodal biometrics have focused on how to fuse the information from different modalities [9]. With this in mind fusion can occur at four levels [7, 11], the most common three of which are depicted in Figure 2:

- **Sensor level**: biometric traits from different sensors are combined into a single block of data. Yang et al. [17] proposed the use of panoramic face images as a mosaic, which produces 3-D surface information, to handle with pose variation. Zhang et al. [19] developed a hybrid mosaic swipe fingerprint scheme to succeed in registering swipe fingerprint frames.

- **Feature level**: signals coming from different sensors are first pre-processed, and feature vectors are extracted separately. Fusion algorithms combine the feature vectors into a single vector that serves as an input for classification. Experiment results for fingerprint feature mosaicing reflect that it performs better than its image counterpart, mainly due to its low memory requirements and low computational complexity [18]. Kumar et al. [8] proposed integration of hand shape and palm print texture at feature level, with extraction of DCT coefficients and shape features respectively.

- **Matching score level**: the vectors are computed separately and an individual matching score is found for each one. The fusion algorithm will work on the individual scores yielding a composite matching score. Ben-Yacoub et al. [1] tested several classifiers like SVM, multi layer perceptron, C4.5 decision tree, Fisher’s linear discriminant, and Bayesian classifier for fusion of face and speech information.

- **Decision level**: Each modality is first pre-classified independently. The final classification is based on the fusion of the outputs of the different modalities. Chatzis et al. [3] proposed five person authentication algorithms based on grey level and shape information of a person’s face and voice features using fuzzy k-mean and fuzzy VQ.

Figure 1: Levels of fusion in a bimodal biometric system: FU - Fusion Module, MM - Matching Module, DM - Decision Module [16].

Successful multimodal recognition systems have been built in different application domains demonstrating the usefulness of information fusion [13]. Dialog Communication Systems developed BioID, commercially available since 1998, is the most widely referred multimodal biometric system, using face, voice, and lip movement to identify people [4].

All the aforementioned systems were developed under the assumption that data was acquired under very constrained conditions such as uniform illumination, user cooperation, controlled noise factors, etc.
new challenge for biometrics arises from trying to circumvent these limitations and developing more robust recognition systems [15]. Multimodal biometric systems working with images acquired under unconstrained settings would be useful both in increasing recognition rates as in evaluating the quality of each trait and performing matching with respect to such values.

3 Preliminary work on iris recognition

Some preliminary work was already performed regarding iris segmentation and recognition on noisy images. The developed segmentation algorithm was presented in [12] and used mutual context information from gradient flow and shortest closed paths to perform robustly with a series of noise factors. The algorithm was tested on the UBIRIS.v2 noisy iris image database [14].

The developed algorithm can be divided in three main modules: the detection of iris centre candidates, by analysing gradient flow divergence; the detection of the best closed contour around each candidate and the discrimination of the best iris centre/contour pair using mutual context information. This discrimination was performed by maximizing a quality factor parameter. Each candidate centre/contour pair was associated with three parameters: gradient divergence around the centre candidate, mean gradient magnitude along the contour and contour circularity. Choosing the candidate pair that presented maximum values for any of these parameters resulted in relatively high misdetection ratios. The quality factor was designed so as to combine all these information sources. A 2.1% misdetection ratio was achieved by discriminating based on the maximization of the quality factor. This result represented a 6.49% improvement over that of the best individual parameter.

A subset of 802 images of the UBIRIS.v2 database was tested, with all the images being manually annotated. The evaluation of the segmentation algorithm was performed by computing distances between the detected and the manually annotated contours. A mean distance of 4.86 ± 2.96 pixels was observed. The distribution of distances is presented on Figure 3. The observed amount of misclassified pixels (pixels falsely considered iris and pixels falsely considered non-iris) was 1.02%. Future work on segmentation will be focused on noise detection so as to avoid the nefarious effect of noisy regions on the recognition process.

Figure 2: Histogram of the distances (in pixels) between the detected contours and the manually annotated ones.

As far as recognition is concerned, the Speeded Up Robust Features (SURF) local keypoint descriptor was used to identify and describe keypoint matches between the segmented iris’ regions. The preliminary results are not very good, with a 39.2% equal error rate (EER). We are currently working on perfecting this algorithm and analysing if the tested database contains enough detail to allow the evaluation of recognition algorithms.

4 Future work

As referred in previous sections the main goal of the proposed work is the development of robust biometric recognition algorithms and further research regarding information fusion to improve recognition rates. However a few other objectives can be presented as more immediate future work, serving as first steps towards the aforementioned goals:

- Creation of a multimodal biometric database: even though several multimodal databases are publicly available [16], we propose the creation and publication of a multimodal database of data acquired under unconstrained settings. Conditions such as motion blur, focus, natural illumination (for image-based traits) or environmental noise and distortion (for voice) are easily modifiable and will be an important part of the proposed database.
- State-of-the-art limitations: the study of the constraints under which the currently available multimodal systems work will constitute an essential step in the development of the proposed work. This step will have a crucial importance in determining the specific needs of the database proposed in the last item.
- Quality assessment: before any attempt is made on the development of unconstrained recognition algorithms a quality assessment module needs to be developed. Such module receives data from the multiple sources of information (i.e. biometric traits) and quantifies if each source presents enough information to be considered valid for the further steps of recognition.

5 Conclusion

A brief overview of multimodal biometrics and the proposed work focused on unconstrained acquisition settings was presented. Future work will focus on the development of robust biometric recognition algorithms, to work with non-ideal data, and information fusion to improve recognition rates.

References

A Biosignal Acquisition Platform based on Arduino and Android

Ana Priscila Alves
AnaPriscila.Alves@lx.it.pt

Hugo Silva
hsilva@lx.it.pt

André Lourenço
alourenco@deect.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
The Arduino is a very popular open-source hardware platform widely used due to its flexibility and usability. It impelled our team to present a low-cost biosignal acquisition system, the Vitalino, based on this platform, and connected to an Android Operating System device, creating a mobile platform for signal acquisition. It can be used to acquire different kinds of biosignals, but we will give special emphasis on electrocardiographic (ECG) signals, acquired with finger-based sensors, due to its integration on a biometric recognition system. The experimental evaluation revealed that our first prototype is capable of ECG signal acquisition with adequate quality, driving the use of this platform for acquiring other types of biosignals. In this paper we describe the proposed platform, with special emphasis on the design principles and functionality. Future work will focus on further developing our hardware, targeting its integration in a prototype system for ECG-based biometric recognition.

1 Introduction
Biosignal acquisition has been a topic of increasingly growing development, since it constitutes the basis for diagnostic systems, and contributes to a better understanding of the body functions. Nowadays, novel applications of biosignals are emerging in areas where they are not traditionally found, such as the use of Electrocardiographic (ECG) signals for biometric purposes. Therefore, our goal was to perform near real-time biosignal acquisition with the Android Operating System (AOS), targeting the development of a mobile biometric system based on ECG signals.

Open source hardware platforms have become increasingly popular, and multiple hardware options are currently available, such as the Google Accessory Development Kit [2], designed to help Android hardware accessory builders and software developers create Android-compatible accessories. In our work we focused on further extending this concept to the context of mobile biomedical applications, which present unique requirements. Although the first use of our acquisition system targets the integration in a biometric platform, allowing real-time recognition, the platform can be extended to numerous other important applications, depending on which signals are to be acquired.

The Biosignal Acquisition Platform - Vitalino - has three main components, namely: Biomedical Sensors, which in our case enable the acquisition of ECG signals; Biosignal Acquisition System, based on the Arduino and used to acquire biosignals and transmit them via Bluetooth wireless to a base station; and a set of Application Programming Interfaces (API) in Java, which communicates with the Arduino board using a custom protocol, controls the acquisition process, allows the access to the collected raw data and enables high-level applications to access both the device and the data. These two parts of the system will be detailed in the next sections.

3 Hardware
The main component of this section is the Arduino. Figure 2 shows the final prototype, with main components: one Arduino Pro Mini (3.3V and 8MHz), directly connected to a Bluetooth Mate module; one Lithium Ion Polymer battery with nominal voltage of 3.7V at 400mAh, as power supply; and one LiPo battery charger. The ECG sensors are connected to one of the Vitalino analog input pins, and its specifications are detailed in [4].

4 Firmware
The main purpose of the firmware is to control the analog and digital acquisition, using a pre-defined sampling rate; all the data acquired is sent to another device via Bluetooth or USB connection. The open-source Arduino programming environment makes it easy to write code and upload it to the I/O board, which is one of the main reasons why this platform was chosen as the base for our system. There are three modes in which the Vitalino can be operating:

a) Live: In this mode, the system continuously samples the physical analog and digital channels, packs all the data, and sends it through the UART. All these tasks should be concluded within a time frame, which is imposed by the sampling rate.

b) Simulated: Although it is similar to what happens in acquisition mode, in this mode, the system will simulate the acquisition, transmitting synthesized signals. These correspond to sinusoidal, square
and sawtooth waves. This way, the communication and interaction between the base station and the device can be tested.

c) Idle: The system disables any mode in which it is in, and stays in stand-by until it receives a command to start the Live or Simulated modes.

The control of the sampling process is performed by the use of interrupts, an external event that stops the running program to execute a special interrupt service routine (ISR). Inside the ISR function, there are some fundamental steps. At first, when in Live mode, 6 analog and 4 digital input pins are read, and their data saved in an array. After concluding this step, all data is sent to the Bluetooth module through the UART. However, a typical 10-bit analog reading on an ATMega MCU takes approximately 111 microseconds to conclude. In our work, we targeted to acquire a new sample at each millisecond, and each sampling includes the resolution and sampling rates up to 1KHz suffice. Additionally, our work and application was developed, containing Start and Stop buttons that calls the corresponding methods on the API. Thus, when the application starts, the base Arduino platform does not have the optimization for power consumption, and sawtooth waves. This way, the communication and interaction between the base station and the device can be tested.

5 Software

5.1 API's

An Application Programming Interface (API) was developed in Java in order to control the Arduino. Its main purpose is to establish a Bluetooth connection, and then start or stop the acquisition, receiving the acquired samples, and configuring the device. This API consists of an abstract class Device, which is subdivided into 2 main subclasses: Bluetooth and Test.

a) Bluetooth: This subclass establishes a Bluetooth connection with the Arduino, and control its operation sending commands that activate the Start or Stop methods. When Start is activated, all data received from the device is saved in a text file for processing.

b) Test: The Test subclass does not communicate with the device, but is used to test the API functionality.

The configuration parameters used in these subclasses, such as baud and sampling rates, are specified in a notation based on JSON (JavaScript Object Notation). It creates an easy, standard, Human-readable and structured way to represent diverse information, and works regardless of the adopted programming language [3]. An example of setup parameters defined using the JSON notation is as follows:

```
{"BaudRate":115200,"Sampling Rate":1000,"MACAddress":"50:01:02:03:04:AA","Mode":"Live"}
```

Focusing on the particular operations that can be performed using the Bluetooth subclass, the main methods are:

1. **Setup**: A JSON Object containing information about Sampling Rate, MAC Address, Acquisition mode (Live or Simulated), and Baud Rate is parsed and, with this information, a Bluetooth connection is established.

2. **Start**: The corresponding start acquisition mode, Live or Simulated, is sent to the system, which will activate its acquisition state.

3. **Acquire**: After activating the acquisition state, this method reads the incoming data, and saves each sample in a text file, using the test application described in the next subsection.

4. **Stop**: When this method is called, the acquisition state is stopped, and the system returns to the idle state.

5.2 Test Application

To test the API functionality and benchmark the device, an Android test application was developed, containing Start and Stop buttons that calls the corresponding methods on the API. Thus, when the application starts, the Bluetooth communication is established using the Setup method, and it remains waiting for a button to be pressed, executing the method Start and Acquire when the button “start” is pressed, and calling Stop method when the button “stop” is pressed.

6 Experimental Evaluation

Tests were performed to the final system, to check the sampling rate and the quality of ECG signals acquired. Therefore, to verify if the Arduino was acquiring at the specified sampling rate initially envisioned for our biometric application of 1KHz, a synthesized square wave with a frequency of 10KHz, duty cycle of 50%, 4 Vpp and offset of Vcc/2 was acquired, and the data was analysed using Matlab. The synthesized wave revealed a square wave, as expected, but after measuring the number of samples in each pulse we verified an average loss of 5 samples per second. However, since our main purpose is to acquire ECG signals, and its bandwidth is approximately 100Hz, a much lower sampling rate of 200Hz can be used for data acquisition, leading to a maximum loss of 1 sample per second, which is negligible for biometric recognition purposes.

![Figure 3: ECG signal acquisition.](image)

In what concerns the ECG signals, the acquisition was performed with a sampling rate of 1KHz. In Figure 3, three isolated heart beats are represented, namely, raw 3(a) and filtered 3(b) with a low-pass Kaiser filter between 2.5 and 30 Hz. We concluded that this system is able to acquire these signals with sufficient quality, being possible to distinguish the different complexes of a characteristic ECG signal, when filtered. Moreover, the battery lifetime was tested, and experimental evaluation has shown that it operates, on average, 6 hours in continuous acquisition.

7 Conclusion and Future Work

We have designed and implemented a first prototype of an ECG acquisition system, applicable to biometric applications. However, this system can also be used to acquire other types of biosignals, becoming a more generic acquisition platform. The higher abstraction level provided by the base Arduino platform does not have the optimization for power consumption or performance that is required for dedicated and specific applications. Still, its versatility and ease-of-use can greatly accelerate the process of prototyping biomedical devices. Furthermore, our work and experimental results have shown that the developed platform performs well enough to be used in applications where just a few channels, 10-bit resolution and sampling rates up to 1KHz suffice.

Future work will be focused on the integration of an external oscillator in the system, separating the execution lines for data acquisition and the data transmission, and increasing the resolution of the system, through an external ADC.

Acknowledgment

This work was partially funded by the IT - Instituto de Telecomunicações under the grant "Android Biometrics Platform" and by the Fundação para a Ciência e Tecnologia (FCT) under grants SFRH/BD/65248/2009 and SFRH/PROTEC/49512/2009 whose support the authors gratefully acknowledge.

References


Identification of benign breasts during mammogram screening

Inês Domingues¹
1 INESC TEC (formerly INESC Porto) and Faculdade de Engenharia, Universidade do Porto
Porto, Portugal
Jaime S. Cardoso¹
http://www.inescporto.pt/~jsc/
Emílio de Azevedo Campos, S.A.
Porto, Portugal
Pedro Cardoso²
http://www.eacampos.pt/

Abstract

A “pre-CAD” system that aims to detect normal mammograms instead of abnormal ones is being designed. This pre-CAD system will work as a “first look” by screening-out normal mammograms, letting the experts focus on “harder” cases. The method consists of two blocks. The first block detects dense breasts and automatically flag them for expert review. Non-dense breasts are sent to the second block where benign cases are identified. Tests made on the INBreast database automatize a significant number of cases, without misclassifying a single malign case.

1 Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females, accounting for 23% of the total cancer cases and 14% of the cancer deaths [5]. It is estimated that, in the United States, 226,870 women will be diagnosed with and 39,510 women will die of cancer of the breast in 2012 [1]. At present there is no known method to prevent breast cancer but early detection and diagnosis increase the chance of cure. Therefore, screening is recommended by the medical community. Screening can, however, be a very tedious and tiring task for specialists. We believe that easy cases can be automatically detected, alleviating the human effort and giving the specialist more time to carefully evaluate more ambiguous cases.

2 State of the art

The first work published in this field was by Sahiner et al in 1996 [9]. The authors investigated the classification of regions of interest (ROI’s) on mammograms as either mass or normal tissue using a convolution neural network (CNN). The input images to the CNN were obtained from the ROI’s using two techniques. The first technique employed averaging and subsampling. The second technique employed texture feature extraction methods applied to small subregions inside the ROI. Features computed over different subregions were arranged as texture images, which were subsequently used as CNN inputs. A data set consisting of 168 ROIs containing biopsy-proven masses and 504 ROI’s containing normal breast tissue was extracted from 168 mammograms by radiologists experienced in mammography. With the best combination of CNN architecture and texture feature parameters, the area under the test ROC curve reached 0.87, which corresponded to a true-positive ratio of 90% at a false positive ratio of 31%.

Babbs and Delp team have published in this field from 1998 to 2007. In their most recent work, a Support Vector Machine (SVM) based method is introduced [2]. Crossed-distribution feature pairs are identified and mapped into new features that can be separated by a zero-hyperplane of the new axis. The probability density functions of the features of normal and abnormal mammograms are then sampled and the local probability difference functions are estimated to enhance the features. From 1,000 ground-truth-known mammograms, 250 normal and 250 abnormal cases, including spiculated lesions, circumscribed masses or microcalcifications, are used for training. The classification results tested with another 250 normal and 250 abnormal sets show testing performances with 90% sensitivity and 89% specificity.

The other group with strong contributions in the field is led by Elshinawy. They noticed that, in general, there are two major types of mammograms according to their tissue type: fatty and dense mammograms. A “pre-CAD” system was designed for detecting only normal mammograms [3]. Mammograms are first separated into two different categories according to their tissue type, and each category is studied individually. A total of 13 features extracted from the Gray Level Co-occurrence Matrices (GLCM). One-class and two-class SVMs were compared. A majority voting approach based on combining the outcome of the classifier of all the blocks that correspond to the Mediolateral oblique (MLO) and Cranio-caudal (CC) views of the same breast was then used. Results showed that separating the mammograms into two disjoint categories reduced the false negative rate in each of the fatty and dense mammograms while keeping the false positive rate as low as possible.

3 Materials and Methods

In this section both the database and the classification methodology used in this work are detailed.

3.1 Database

INBreast [6] is a recently proposed database that, due to its contour quality, we believe will become a benchmark in the field. It consists of 205 breasts with MLO and CC views (making a total of 410 images). Images area ranges from 8 to 14 × 10⁶ pixels² (average area of 11 × 10⁶ pixels²) with an average height of 2.9 × 10³ and an average width of 3.6 × 10⁵. Besides contour information, INBreast also provides ground truth information on breast density (classified according to the ACR standard) and malignancy (in the BI-RADS standard).

Breast density classes were binarized by considering ACR I and II as non-dense and ACR III and IV as dense. In this way, 283 images are non-dense, while the remaining 127 are considered dense. Malignancy was also binarized by joining BI-RADS classes 1, 2 and 3 (by assuming they correspond to benign findings) and similarly classes 4, 5 and 6 (assuming they are malign). In total, 310 images were considered benign and 100 malign.

3.2 Benign breasts identification

Each breast (MLO and CC views) is fed into system whose flow chart can be seen in Figure 1.

Figure 1: General block diagram.

The density classification block only uses the CC views and outputs a binary result. As density is strongly associated with breast cancer and decreases the sensitivity of automatic systems [10], breasts classified as dense are directly forwarded to a specialist. Otherwise, the breast goes through the Benign breasts identification block. In this stage, both views are analysed and again a binary output determines if the breast is benign or suspicious. If benign, the general, country specific, routine screening is advised, however if suspicious, no automatic decision is made and the breast image is passed to a specialist. The next sections detail the classification blocks.
3.2.1 Density classification

In this block, only the CC view is used. Intensities are first normalized by histogram stretching to the range $[0,1]$. Then, a histogram analysis (by comparing the average intensity level of the right half with the left half) is made to mirror, if necessary, the image so that the nipple is pointing to the right side. The breast region is detected by using the Otsu threshold [8]. Some features are then extracted from the breast region. After some studies, it was observed that the most relevant characteristics are the 32 bin histogram (from which only 8 bins were kept, namely bins 8, 9, 10, 11, 12, 16, 17, and 24) and a SVD feature as described in [7], making a total of 9 features. These features are normalized to have zero mean and unit variance. A SVM classifier is then created with an RBF kernel and a grid search is performed to select the optimal values for $C$ and $\gamma$.

3.2.2 Benign breasts identification

Non-dense breasts are passed to the benign breasts identification block. Pre-processing steps (histogram stretching, mirroring and breast detection) are performed as described above. The image is then divided into blocks of $512 \times 512$ (around 40 blocks for each image) and, for each block (with at least 40% of breast tissue), 15 features are extracted. GLCM matrices with distance 1 and angles $0^\circ$, $45^\circ$, $90^\circ$ and $135^\circ$ are constructed. From each one of the 4 matrices, 5 properties are extracted: Contrast, Correlation, Energy, Entropy and Homogeneity. The mean and standard deviation of these 5 properties are then calculated. The last feature is derived from the Local Binary Pattern (LBP) image by computing its contrast value. The average values of each one of the above features over all the image blocks is taken as the feature vector. The final breast feature vector of length 30 corresponds to the concatenation of the features for the two views. Again, the features are normalized and a SVM classifier is created as before.

4 Results

The INBreast database was divided into train and test sets (75%/25%). The density block classifier presents an accuracy of 50% and False Negative rate $FN = 6\%$. Some results can be seen in Table 1.

![Figure 2: Benign breast recommended for routine screening.](image)

<table>
<thead>
<tr>
<th>Dense correctly classified</th>
<th>Dense incorrectly classified</th>
<th>Non-dense correctly classified</th>
<th>Non-dense incorrectly classified</th>
</tr>
</thead>
</table>

The benign breasts identification classifier was trained with 46 breasts (26 benign and 23 malign) classified as non-dense by the previous block. The performance on the test set of 69 breasts is: True Positive rate $TP = 12\%$, True Negative rate $TN = 28\%$, False Positive rate $FP = 61\%$ and False Negative rate $FN = 0\%$. Some results can be seen in Figures 2, 3 and 4.

![Figure 3: Benign breast sent to be reviewed by an expert.](image)

![Figure 4: Malign breast sent to be reviewed by an expert.](image)

5 Conclusion

We have presented a fully automatic benign breast identification system. To the best of our knowledge this is the first work that uses an automatic density classification block to filter images before passing them to the Benign breasts identification classifier. Future work consists of improving both classifiers and in testing the system in a real screening environment.

6 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 Tecnologia (Portuguese Foundation for Science and Technology) within projects FCOMP-01-0124-FEDER-022701, PTDC/SAU-ENB/114951/2009 and SFRH/BD/70713/2010.

References

Color feature selection for unconstrained iris recognition

Ana F. Sequeira
afps@inescporto.pt
Jaime S. Cardoso
jaime.cardoso@inescporto.pt
Samaneh Khoshrou
skhoshrou@inescporto.pt
INESC TEC and Faculdade de Engenharia
Universidade do Porto
Porto, Portugal

Abstract
In recent years many authors have recognized that the path forward, regarding iris recognition, is the development of iris recognition systems that can work independently of the conditions under which iris images are acquired. Recent works have tried to achieve robust and unconstrained iris recognition in order to develop real-world applicable methods. In this work, the problem of unconstrained iris recognition is referred and some results of an approach to the problem of fusing color information to enhance the performance of an iris authentication system are briefly presented.

1 Introduction
Reliable automatic recognition of persons has long been an attractive goal. In most of all daily activities, personal identification plays an important role. The most traditional techniques to achieve this goal can be divided in two kinds: knowledge-based and token-based. In one hand, token-based approaches take advantage of a personal item, like a passport, driver’s license, ID card, credit card or a simple set of keys, on the other hand, knowledge-based approaches, are based on something the user knows that, theoretically, are not accessible to others, such as passwords or personal identification numbers. These approaches present obvious disadvantages: tokens may be lost, stolen, forgotten or misplaced, while passwords can easily be forgotten by a valid user or guessed by an unauthorized one [5]. In fact, all of these approaches stumble upon an obvious problem: any piece of material or knowledge can be fraudulently acquired.

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.

Several biological traits in humans show a considerable inter-individual variability: fingerprints and palmprints, the shape of the ears, the pattern of the iris, among others. Biometrics works by recognizing patterns within these biological traits, unique to each individual, to increase the reliability of recognition. The growing need for reliability and robustness, arise some expectations and become the focal points of attention when someone is trying to develop a new system based on a specific trait: universality, uniqueness, collectability, permanence [5].

Iris patterns become interesting as an alternative approach to reliable visual recognition of persons when imaging can be done at distances of less than a meter, and especially when there is a need to search very large databases without incurring any false matches despite a high number of possibilities. Although small (11 mm) and sometimes problematic to image, the iris has the great mathematical advantage that its pattern variability among different persons is enormous [2]. The iris presents itself as a leading candidate to become the standard biometric trait: it is universal, the variability is huge which assures the uniqueness for each individual, apart from being an organ easily accessible and very difficult to modify.

2 (Unconstrained) iris recognition system
The pioneer works in iris recognition set the basis of the typical iris recognition system architecture: segmentation, normalization, feature extraction and matching [3]. For an overview of the system see Fig. 1.

Currently, there are several biometric systems based on iris recognition with excellent rates of success. However, this performance is due to the constrained conditions under which iris data is acquired (infra-red illumination of the eye, user collaboration, controlled distance from the camera, etc.). Some implicit or explicit assumptions about the acquisition process are no longer valid in unconstrained acquisition scenarios. Therefore, some of the promising results reported in the literature must be taken with caution and reassessed under these new, more challenging, conditions.

The new challenges for iris biometric systems arise when the attempt is made to perform iris recognition without user cooperation or under less ideal conditions. Some typical unconstrained scenarios in applications may be iris recognition with mobile phones for security such as in airports, in military applications [13] or in bank accounts [1]. It has been recognized that the path forward is the development of algorithms that can work independently of subject collaboration and proper near infrared illumination conditions, in order to achieve robust (i.e. accurate even with noisy images) and unconstrained (i.e. accurate for several sets of acquisition conditions: distance, movement, illumination, etc.) iris recognition and, in this way, become a real-world applicable method [7, 11].

3 Color feature selection for an unconstrained iris recognition system
In this work, we consider the problem of fusing multiple color channels to enhance the performance of an iris authentication system. The verification process is based on open-source implementation made by Masek [6] (available in open source1 and tested with near infrared images).

3.1 Fusion Methods
Multiple expert fusion aims to make use of many different classifiers to improve the classification performance. Different color channels show different performance in various applications. So, it is expected that better performance could be obtained by fusing classifiers based on different color channels. Among the possibilities, we choose two simple approaches, “Averaging” and “Product”; in addition to a sequential search approach. An adopted version of the sequential search approach [10], “Plus L and take away R” works based on finding the best “L” features in the beginning and then, try to find “R” worse features from our last optimum subset. This method finds the best set leading to the best result generally. For all the cases we consider the same features [12].

3.2 Method
We convert the images from RGB to other color spaces. It is common to describe color as a set of three primary colors (Red, Green and Blue) but there are different color spaces that can be used. We use opponent color channels given by $RG = R - G$, $RB = R - B$, $GB = G - B$, Intensity and also HSV (for more details see [4]).

An histogram equalization is applied for photometric normalization of images. The segmentation of the iris region was done manually (due to the difficulty of applying the segmentation part of the method of Masek [6] to the noisy images of the database chosen (UBIRIS v2 [9]) and consisted of selecting three points: the center of pupil and iris, one point in the border of pupil and one point in the border of iris. In this process two “major” assumptions were made: the coincidence of the centers of pupil and iris and the circular shape of both regions 2. Also the noise mask that should be obtained with the information of the occluded regions was considered to be empty, so the iris image used in the posterior process had some noise that was considered as iris regions. After the manual segmentation we used Masek’s code for normalization of the iris image, for the extraction of features and matching. Once the raw scores in different coloration we used Masek’s code for normalization of the iris image, for the extraction of features and matching. Once the raw scores in different color spaces were extracted then we could start the fusion part of the method. We worked on the prediction using a support vector machine (SVM) and tried to apply sequential search method on this new set of features. The sequential search algorithm, “Plus 2 and Take away 1” (“+2 – 1”), was applied for selecting an optimum subset of the color channels. The system uses the prediction of SVM, which is trained based on hamming distances of client and impostor classes, as input. The selection procedure keeps adding or taking away features (color channels in our case) until the best evaluation performance is achieved. The optimum set found is applied to evaluate the performance of the system in the test step. Two different approaches (averaging and product) are also employed to give a critical view on differences between various fusion methods.

4 Experimental Setup

4.1 Dataset

We used 10 images of each of the 40 different subjects selected from the UBIRIS.v2 database [9]. The major purpose of the UBIRIS.v2 database is to constitute a new tool to evaluate the feasibility of visible wavelength iris recognition under far from ideal imaging conditions.

4.2 Experimental Results and Discussion

In the following tables, for each different color spaces, the results are shown using: FARE, FRRE and TERE (False Acceptance, False Rejection and Total Error, respectively, Rate for Evaluation); FART, FRRT and TERT (False Acceptance, False Rejection and Total Error, respectively, Rate for Training). Also the boundary surface for SVM was displaced for minimizing the difference between the FAR and FRR for training (Equal Error Rate, EER).

The results of Table 1 were obtained using a SVM considering EER in the individual color channels shown in the table. The best performance is obtained in intensity color space. The results of Table 2 were obtained using SVM considering EER in three different fusion methods. The best performance is obtained for “Plus 2 and take away 1” method.

In this experiments, the results obtained by adapted “Plus 2 and take away 1” algorithm outperforms “Averaging” and “Product”. The best result in individual subspaces was 35.5% for intensity space (see Table 1) and for the fusion methods we had obtained a result of 30.8% for “+2 – 1” method. So, as expected the results improved considerably.

4.3 Conclusions

We addressed the problem of using colour information in an iris recognition system. We concluded that, fusing the colour information improves our system. Even using a more intelligent method, sequential search plus SVM method outperforms the blind fusion algorithms such as product and averaging.

This preliminary results encourage further work, but one important aspect to take in account is to use a segmentation method that is suitable for an unconstrained scenario in order to overcome the difficulty observed in applying Masek’s implementation. Also, more color channels are too be tested.

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-022701 and PhD grant with reference SFRH/BD/74263/2010.

References


Table 1: Results (%) using SVM for different color spaces considering EER

<table>
<thead>
<tr>
<th>R</th>
<th>G</th>
<th>B</th>
<th>Intensity</th>
<th>H</th>
<th>S</th>
<th>V</th>
<th>RG</th>
<th>RB</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>FARE</td>
<td>24</td>
<td>22.5</td>
<td>30.2</td>
<td>19.9</td>
<td>45.8</td>
<td>31.9</td>
<td>23.8</td>
<td>34.5</td>
<td>26.2</td>
</tr>
<tr>
<td>FRRE</td>
<td>24</td>
<td>22.7</td>
<td>30.6</td>
<td>19.9</td>
<td>44.9</td>
<td>31.6</td>
<td>23.6</td>
<td>33.9</td>
<td>26.7</td>
</tr>
<tr>
<td>TERE</td>
<td>48</td>
<td>45.2</td>
<td>60.8</td>
<td>39.8</td>
<td>90.7</td>
<td>63.5</td>
<td>47.4</td>
<td>68.4</td>
<td>52.5</td>
</tr>
<tr>
<td>FART</td>
<td>17.8</td>
<td>20.3</td>
<td>27.5</td>
<td>17.3</td>
<td>45.1</td>
<td>33.1</td>
<td>16.9</td>
<td>30.8</td>
<td>22.3</td>
</tr>
<tr>
<td>FRRT</td>
<td>22.7</td>
<td>19.7</td>
<td>27.4</td>
<td>18.2</td>
<td>39.7</td>
<td>22.6</td>
<td>24.8</td>
<td>35.8</td>
<td>28.6</td>
</tr>
<tr>
<td>TERT</td>
<td>40.5</td>
<td>40</td>
<td>54.9</td>
<td>35.5</td>
<td>84.8</td>
<td>55.7</td>
<td>40.9</td>
<td>66.6</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 2: Results (%) for Fusion methods using SVM considering EER

<table>
<thead>
<tr>
<th>R</th>
<th>G</th>
<th>B</th>
<th>Intensity</th>
<th>H</th>
<th>S</th>
<th>V</th>
<th>RG</th>
<th>RB</th>
<th>GB</th>
</tr>
</thead>
<tbody>
<tr>
<td>+2</td>
<td>-1</td>
<td>Averaging</td>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FARE</td>
<td>16.7</td>
<td>19.4</td>
<td>19.8</td>
<td>19.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRRE</td>
<td>16.9</td>
<td>19.4</td>
<td>19.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERE</td>
<td>33.6</td>
<td>38.8</td>
<td>39.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FART</td>
<td>14.6</td>
<td>15.7</td>
<td>16.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRRT</td>
<td>16.2</td>
<td>18.1</td>
<td>18.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERT</td>
<td>30.8</td>
<td>33.8</td>
<td>34.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Automatic assessment of Leishmania infection indexes of in vitro macrophage cell cultures

Marco Marques\(^1,2\); Pedro Leal\(^1,2\); Luís Ferro\(^1,2\); Susana Romão\(^3\); Tânia Cruz\(^1\); Ana M. Tomás\(^3\); Helena Castro\(^1\); Pedro Quelhas\(^2\)

\(^1\)Universidade do Porto, Faculdade de Engenharia, Departamento de Engenharia Química, Portugal
\(^2\)INEB - Instituto de Engenharia Biomédica, Portugal
\(^3\)IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Portugal

Abstract

Evaluation of 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 particular case of Leishmania infantum, a unicellular parasite that parasitizes macrophages, infection indexes are usually determined either by visual inspection of cells directly under the microscope or by counting digital images using appropriate software [1]. We propose a fully automated method for automatic evaluation of parasite infection indexes through the segmentation of individual macrophages nucleus and cytoplasm, as well as the segmentation and co-localization of the parasites. To perform such analysis with robustness and increased performance we propose the use of local image filters tuned to the specific size of the objects to detect, in conjunction with image segmentation approaches. The objects’ size estimation is then improved through a learning feedback loop. Cytoplasm is detected by seeded watershed segmentation. Our approach obtains, for 86 images from 4 experiments, an average parasite infection index evaluation error of 2.3%.

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 cultures macrophages infected with the parasite. Under this in vitro setting, macrophage infection indexes are used as a measure of parasite growth and/or 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, either by visual inspection of cells under the microscope [2]. Macrophages and parasites are labeled with immunofluorescent chromophores that label macrophages’ nuclei in red and parasites in green.

Most approaches for cell detection in fluorescence microscopy images are based on automatic image segmentation [3, 5, 6]. The specific methods used range from image thresholding or more advanced methods such as mean-shift clustering [4]. In the case of touching cells watershed transform is widely used [5, 6]. While these methods produce reasonable results in good quality images the parameters are unintuitive and nonrobust to changes in image quality.

We propose an approach based on Difference-of-Gaussians (DoG) local filtering to enhance nuclei and parasite locations, where DoG size is initialized by measuring the objects in the segmentation of the original image. This is performed using a feedback loop where Otsu thresholding is used to segment the nuclei/parasites and estimate their size [7]. We then use the newly estimated size for the DoG filter parametrization. The process is repeated until no change in estimated size is detected. This novel adaptation feedback loop leads to high robustness to errors in initial estimated size. Nuclei based seeded watershed is used to detect macrophage cytoplasmic area and parasite infection index is evaluated based on parasite/cytoplasm overlap.

2 Image Dataset

To evaluate our approach we gathered 86 in vitro color images of monolayer macrophages cultures obtained by fluorescence microscopy, of macrophages infected with L. infantum. Each image color channel has information on the spatial location and shape of macrophage’s cell nuclei (B channel), cytoplasm (R channel) and L. infantum parasites (G channel).

Figure 1: Examples of images of L. infantum infected monolayer macrophages cell cultures obtained by fluorescence microscopy (cropped in height).

3 Methods

3.1 Pre-Processing

Pre-processing for each channel starts with Gaussian smoothing filtering (σ = 2), which has the objective of removing noise, then an adaptive image histogram equalization is performed to remove image illumination irregularities.

3.2 Nuclei and parasite detection with size selection

Cell detection achieved by segmentation is difficult to tune for the specific image conditions and is not robust to varying image conditions. We approach the problem differently where the parameters in our approach are intuitive, the size of the objects to detect, and robustness is obtained using a self regulating feedback loop.

Using the pre-processed images from our dataset we implement an approach that is based on the scale selection properties of the difference-of-Gaussians (DoG) filter. This process, in our case, enhances both cell nuclei and parasites prior to segmentation. After the application of the DoG filter to our input image, we use Otsu automatic threshold segmentation to obtain the detection result [7].

To overcome the need to know the size of the objects, DoG parameters are tuned based on the data. Starting with the 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.

Figure 3 shows the iteration of the segmentation of nuclei and the re-estimation of the DoG size from different initial guesses. As can be observed the initial segmentation (Figure 3 - Otsu:1) is not accurate. However, the feedback loop continues improving parameters leading to a better segmentation. (Figure 3 - Final Seg.). The same approach is used in the case of parasite detection, with a different image channel as input and a suitable initial size guess.

3.3 Cytoplasm segmentation

The only prior knowledge we can use for cytoplasm segmentation is that for each cell its nucleus is inside the cytoplasm. Given the location of nuclei (B channel) inside the cytoplasm area (R channel) we use seeded watershed on the Red image channel as our cytoplasm segmentation approach, in which the nuclei location is the seed for the watershed. From the watershed result we then subtract the background, obtained using Otsu
This dataset comprises a total of 11342 cells and 6729 parasites, according to the image and each parasite which is infecting that specific macrophage. The expert researcher annotated both the location of each nuclei in the image and each parasite which is infecting that specific macrophage. More importantly, locations for nuclei and parasites are not centered and sometimes do not overlap with the respective object in the image. As can be seen in Figure 4 many user given annotations are off center. These results may have been inflated due to the lack of precision of manual annotated results for infection level estimation. 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 experiments.

4 Results

Each image was annotated by hand by an expert researcher using the CellNote annotation software implemented at the IBMC, Porto [1]. For each image the expert researcher annotated both the location of each nuclei in the image and each parasite which is infecting that specific macrophage. This dataset comprises a total of 11342 cells and 6729 parasites, according to the annotation.

For an objective measurement of our approach’s performance we calculated several measures of performance. For macrophage nuclei and parasite detection we calculated the detection error (DE), presenting also the values for false positives (FP) and false negatives (FN) to distinguish type I and type II errors. As the final objective of our approach is to estimate the infection level in the cell culture we calculate the infected cell count error (ICE) and the number of parasites infecting each cell error (NPIE). ICE is evaluated as the error in detection of infection of cells when compared with the annotated groundtruth and NPIE is the difference between the number of parasites for each infected cell between automatic and annotated counts. Table 1 shows the evaluation numbers for our approach, for each of the 4 experiments as well as overall averages.

In terms of image segmentation based detection error we obtain an average error of 29% and 19% for cell nuclei and parasite detection respectively. These results may have been inflated due to the lack of precision of the annotating experts, due to their objective of evaluating infection level and not cell/parasite location. As can be seen in Figure 4 many user given locations for nuclei and parasites are not centered and sometimes do not even overlap with the respective object in the image. More importantly, in the evaluation of the parasite infection index our approach is within a maximum error of 2.3% from the manual annotation. Figure 4 shows the manual and automatic results.

5 Conclusions

We proposed an automatic approach for the evaluation of *L. infantum* parasite infection index on an in vitro macrophage’s cell culture, based on fluorescence microscopy digital images. Using a learning feedback loop for the estimation of cell nuclei and parasite size we were able to increase robustness of the applied segmentation methods.

Results obtained on 86 images show that our approach is 2.3% of manual annotated results for infection level estimation. 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.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Nuclei</th>
<th>Parasites</th>
<th>Infection Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>DE</td>
<td>FP</td>
<td>FN</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>159</td>
<td>95</td>
</tr>
<tr>
<td>2</td>
<td>35</td>
<td>51</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>total</td>
<td>86</td>
<td>39</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1: Performance results: Detection error (DE), false positives (FP) and false negatives (FN) average values over all images are presented for the detection of cell nuclei and parasites (standard deviation across all images in parenthesis). Parasite infection level estimation was also evaluated by the infected cells’ counting error (ICE) and the number of parasites infecting each cell error (NPIE). NC is the average number of cells according to the annotation.

References


Local interest detector based cancer cell mobility and morphology joint analysis

Tiago Esteves\textsuperscript{1,2}
dee11017@fe.up.pt

Maria José Oliveira\textsuperscript{1,3}
mariojo@ineb.up.pt

Pedro Quelhas\textsuperscript{1,2}
quelhas@fe.up.pt

Abstract
The automatic analysis of cell mobility has gained increasing relevance given the enormous amount of data that biology researchers have currently to analyze [5]. During cancer cell invasion and metastasis, cancer cells acquire, in response to the stimuli to which they are exposed, distinct mobility or motility capabilities. While mobility consists on cell migration and movement along a given surface, motility corresponds to cell morphological alterations towards a given stimulus and does not require necessarily cell movement. By evaluating cell motility/mobility towards distinct stimuli or inhibitors, more efficient therapies can be then designed [3]. To evaluate cancer cell’s mobility, biologists establish in vitro invasion assays with cancer cells seeded on native surfaces or on surfaces coated with extracellular matrix components, recording time-lapse brightfield microscopy images [3]. In such analysis only through the use of quantitative automatic analysis tools it is possible to gather evidence to firmly support biological findings. However, most automatic cell tracking approaches are based on non-robust image segmentation approaches and do not take into account cell morphological alterations that might occur along the tracking process. Herein, we develop a new tracking approach that incorporate cell morphology modeling for higher performance and adaptability.

1 Introduction
The analysis of cancer cell’s mobility and morphology is fundamental for the understanding of more complex cellular responses such as cell invasion and for the development of possible therapies for its regulation. Usually, for studying cancer cell motility and mobility, cancer cells are placed on top of native surfaces or of extracellular matrix-coated surfaces and a time lapse video is collected along their motion as interaction with the substrate. A crop of a typical image can be seen in Figure 1. While all cancer cells are similar at the beginning of the experiment, they behave differently during the time-lapse movie, changing their morphology and reducing their mobility (Figure 1 - dotted green ellipse) or remaining mobile, spreading along the surface (Figure 1 - dotted red ellipse). These results suggest that the shape and morphology of each cell is directly related to both surface interaction/adherence capability but also the mobility of each cell. Additionally, cells eccentricity gives a clue of their direction of motion as they tend to move more in the direction of elongation. We propose to use this relation between shape and mobility to improve the cell tracking process.

However, most automatic analysis tools model mobility and morphology independently and are based on non-robust image segmentation approaches. Herein, we introduce a new methodology which is based on local interest point detection and analysis of the Hessian matrix to improve cell tracking. Contrary to most cell tracking and analysis software we do not use segmentation, avoiding the problems of variability of image properties, over and under-segmentation and cell clustering [4]. Instead, we rely on local filters which respond directly to specific shapes in the image, more likely to be cells. This approach has already shown to have high performance in previous works [1,3].

1.1 Cell detection
Automated cell tracking in microscopy image is, in most cases, based on a detection-association methodology, where detection is obtained using automatic image segmentation [1]. However, image segmentation is a fundamental and difficult problem in computer vision.

We base our approach on a more robust detection approach based on the scale-normalized Laplacian of Gaussians (LoG) filter to enhance the image’s blob like structure which correspond to cell locations (figure 2 - a). Given an input image \(I(x,y)\), the gaussian scale space representation at a certain scale \(t\) is:

\[
L(x,y,t) = g(x,y,t) * I(x,y), \quad \text{where} \quad g(x,y,t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}}. \quad (1)
\]

The LoG operator is defined as:

\[
\nabla^2 L(x,y,t) = L_{xx}(x,y,t) + L_{yy}(x,y,t), \quad (2)
\]

where \(L_{xx}\) and \(L_{yy}\) are the second derivatives in \(x\) and \(y\) respectively. We perform detection of cells by detecting local minima of LoG response in the input image (figure 2 - b), over a range of different \(t\) values for all spatial locations. The detected minimum enables us to estimate the position and radius \((r = 1.5 \times t)\) of cells in the image (figure 2 - c) [2].

To further refine the shape of cell’s we use the Hessian matrix to estimate the eccentricity of the cell’s [2]:

\[
H(x,y,t) = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix}. \quad (3)
\]

From the eigenvalue of the Hessian matrix we obtain the orientation and eccentricity values for the respective ellipsoid shape approximation (figure 2 - d). The eccentricity information enables to estimate the cell’s direction and type of motion.

1.2 Cell tracking
From the cell detection the tracking of cells is performed using a detection-association approach, based on the cell detections distance between consecutive frames. To improve tracking this approach relies also on the cell’s eccentricity, as this is a good clue for the direction of motion, giving preference to associations in the direction of most likely motion (arrows on figure 1). Additionally, the association of cells across frames is also improved by assuming that cell’s shape evolution is not abrupt (figure 3). The applied distance formula was:

\[
D(x_1, x_2) = |x_1 - x_2|^T H(x_1)^{-1} |x_1 - x_2| + k ||\nabla^2 L(x_1) - \nabla^2 L(x_2)||, \quad (4)
\]

where \(x_1\) correspond to the \((x,y,\theta)\) coordinates and scale of the cell being tracked in the current frame and \(x_2\) correspond to a candidate for correspondence in the next frame. The eccentricity information is considered.
Figure 2: Different steps during cell detection: a) Brightfield frame with cancer cells; b) LoG response; c) Detections overlaid in the image; d) Cells detections shown as ellipses based on Hessian estimated eccentricity.

Figure 3: Tracking results showing the association of cells across 5 frames.

Table 1: Tracking performance evaluation results, considering associations using only Euclidean distance and our approach. The results are presented in percentage values.

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>Coverage</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eucl. Dist</td>
<td>30</td>
<td>62</td>
<td>80</td>
<td>69</td>
<td>74</td>
</tr>
<tr>
<td>Our Dist (k = 0)</td>
<td>31</td>
<td>63</td>
<td>81</td>
<td>64</td>
<td>76</td>
</tr>
<tr>
<td>Our Dist (k = 1)</td>
<td>31</td>
<td>61</td>
<td>81</td>
<td>64</td>
<td>77</td>
</tr>
<tr>
<td>Our Dist (k = 3)</td>
<td>27</td>
<td>60</td>
<td>86</td>
<td>55</td>
<td>79</td>
</tr>
<tr>
<td>Our Dist (k = 5)</td>
<td>24</td>
<td>54</td>
<td>83</td>
<td>51</td>
<td>75</td>
</tr>
<tr>
<td>Our Dist (k = 10)</td>
<td>25</td>
<td>48</td>
<td>82</td>
<td>36</td>
<td>74</td>
</tr>
</tbody>
</table>

as we incorporate the hessian matrix ($H$) in the distance formula. The second part of the distance formula is related with the LoG filter’s response magnitude of the cells under analysis, where minimum changes in these values favour the association between detected cells. The $k$ parameter is a distance weighting value that controls the influence of this information in the distance formula.

1.3 Results

We evaluate our approach on our test data, 21 frames containing 81 cells manually annotated. From the obtained results (table 1) we can observe that by using our local eccentricity based distance we were able to improve the tracking by association results, from 74% to 76% (considering $k = 0$) which corresponds to 34 less errors, improving both the amount of frames in which a cell is correctly tracked (from 30% to 31%) and the amount of cells that are correctly tracked in our data (from 62% to 63%). By adding the LoG response to the distance formula, by changing the $k$ parameter to 3, the accuracy value over the full movie where higher (79%) which means that when compared to the result obtained with the Euclidean distance more 85 cells where correctly identified.

In order to allow biology researchers to use our work in their experimental analysis tasks we are developing an easy to use software to perform cell detection and tracking.

The current software capabilities are:

- Open and play videos of the performed experiments;
- Perform single cell tracking through time;
- Perform multiple cell tracking through time;
- Perform cell mobility analysis.

The software user interface is visible in figure 4 and it will be improved according to the specifications of the biology researchers.

2 Conclusion

We concluded that by using local anisotropy clues given by the Hessian matrix it was possible to correctly estimate the cells eccentricity and to use that information to improve tracking results. The results were improved further by using the LoG filter’s response magnitude values as we were able to track more cells correctly for the full movie’s length.

3 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.

Tiago Esteves is recipient of SFRH/BD/80508/2011 by FCT.

References

Vital Analysis: a framework for annotating physiological signals of first responders in action

P. Gomes\footnote{Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal}
ptmgomes@dcc.fc.up.pt

M. Kaiseler\footnote{Instituto de Telecomunicações and Laboratório de Reabilitação Psicossocial, FEUP/FPCEUP, Portugal}
mkaiseler@fpce.up.pt

C. Queirós\footnote{Laboratório de Reabilitação Psicossocial, FPCEUP/ESTSIPP, Portugal}
cqueiros@fpce.up.pt

B. Lopes\footnote{Laboratório de Reabilitação Psicossocial, FPCEUP/ESTSIPP, Portugal}
brunolopes@dcc.fc.up.pt

M. Coimbra\footnote{Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Portugal}
mcoimbra@dcc.fc.up.pt


cmeimbraga@dcc.fc.up.pt

Abstract

First responders such as firefighters are exposed to extreme stress and fatigue situations during their work routines. It is thus desirable to monitor their health using wearable sensing but this is a complex and still unsolved research challenge that requires large amounts of properly annotated physiological signals data. In this work we show that the information gathered by our Vital Analysis Framework can support the annotation of these vital signals with the stress levels perceived by the target user, confirmed by the analysis of more than 4600 hours of data collected from real firefighters in action, including 717 answers to event questionnaires from a total of 454 different events.

1 INTRODUCTION

A first responder is a person trained to intervene in emergency situations in order to help the general population. More specifically, in the case of a firefighter (FF), he is trained to perform under dangerous and high-pressure situations that are both physically and psychologically demanding. These demands can result in problems caused by the long time exposure to several traumatic, non-traumatic, and organizational stressors over time [1], which are known to be associated with psychological and physical illness [6]. This motivates the need for new systems and technologies capable of monitoring, in real time, the physiological signals and the behaviour patterns of these professionals.

With the goal of providing a first response monitoring system on critical emergency scenarios, the “Vital Responder” project was created. It is based on a wearable shirt (Vital Jacket\textsuperscript{©}, in Fig. 1) that is capable to continuously collect electrocardiogram, accelerometers and GPS signals in real time. However, signal processing and machine learning research, aiming to combine these signals into an estimation of an individual’s stress levels, requires the data to be complemented by adequate annotation that can contextualize it. Did an individual’s heart rate rise because he was stressed? Or did he simply start running? For this purpose, we have created the Vital Analysis Framework (VA), which is a smartphone based solution capable of annotating physiological signals of FF in action with both context (details about the event the FF was in) and perceived psychological stress levels (retrieved from the analysis of psychological questionnaires). In this paper, we will focus on the analysis of the quality of our dataset.

Section 2 will describe the Vital Analysis Framework workflow. In Section 3 we will present some results followed by conclusions in Section 4.

2 VITAL ANALYSIS FRAMEWORK

Given the nature of a firefighter’s work our framework needs to collect a reasonable variety of information which will enable us to label and contextualize the physiological data gathered. This includes not only information about official events encountered, but also: information about work being done inside the fire station; information about anomalies or difficulties that appear during an event; and their subjective self assessment of their levels of stress and fatigue in the beginning of the day, after such events, and in the end of a work day. Therefore we have designed three different annotations methodologies.

2.1 Stress Annotation Methodology

Most self-report measures aiming to access stress levels include questions related with physical and cognitive symptoms of stress. Thus, following this principle, our measures of stress included 4 questions related with physical and 4 questions related with cognitive aspects, used previously in validated stress questionnaires [4]. Participants were asked to rate each item on a free scale ranging from "0" to "4", where a rating of "0" represents not felt at all, and a rating of "4" extremely felt. These questions were fulfilled at the beginning and end of the day, aiming to evaluate whether there were alterations in stress symptoms experienced. Furthermore, end of the day and beginning of the day stress symptoms mean scores will be subtracted in order to accomplish an overall mean score, symbolizing accumulated stress symptoms over the day. Internal consistency of the 8 questions was calculated, using Cronbach’s alphas, resulting in 0.93. Additionally, another question was fulfilled after each event, indicating stress appraisal of the event. Participants were asked to rate how they appraised each stressful event, on a free scale ranging from "0" to "4", where a rating of "0" was not at all stressful, and a rating of "4" was extremely stressful.

2.2 Event-driven Annotation Methodology

The Event-driven Annotation gives us the possibility to detail an event by dividing it into several predefined stages, allowing us to evaluate and quantify the collected physiological signals differently for each one. These predefined stages are the basic stages for every single event, and are usually consecutive. A normal event starts with as emergency call, followed by the trip to the event, the event itself, and finally the return trip to the headquarters. Nevertheless special occasions can occur such as: a high priority call during any period of other event; or the cancellation of an event.

2.3 Voice Annotation Methodology

Motivated by the unpredictability of a firefighter’s job we have designed the Voice Annotation Methodology. This methodology will be our “break the glass” mechanism, allowing that at least one of the firefighters can report unexpected activities that happen during an event, or add valuable psychological information, allowing for rich and expressive contributions.

Figure 1: Images of a version of the Vital Jacket\textsuperscript{©}, specially made for the firefighters, and the Vital Analysis framework running on a smartphone.
This methodology is also used to allow the user to add annotation outside events and to enrich the data gathered using the previous methodologies.

3 RESULTS

The Dataset compiled for this study was collected from 12 firefighters, with a mean age of 37.8, between July, 2011 and January, 2012. We have collected more than 4600 hours of data from which we have retrieved a total of 319 days with 717 answers to the event questionnaires.

3.1 Usability Evaluation

To evaluate the usability of our framework we have gathered the official information about the events and compared it with our annotations. The measures chosen were: the percentage of real events that were annotated; the percentage of annotations that were done correctly, in which a correct (good) annotation is one that has all the stages of an event, with a time difference between them above 1 minute; the percentage of annotated events with audio annotation; and the percentage of event questionnaires with audio annotation.

Table 1: Results from the usability evaluation. We have analysed all events, as well as when only a specific amount of FF where present, and when more than a specific amount was present

<table>
<thead>
<tr>
<th>Event categories</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fire</td>
<td>1,23</td>
<td>1,002</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Accident</td>
<td>1,40</td>
<td>0,857</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Pre-hospital assistance</td>
<td>1,11</td>
<td>0,813</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Legal conflict</td>
<td>0,80</td>
<td>0,837</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Services</td>
<td>1,01</td>
<td>1,033</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Activities</td>
<td>0,77</td>
<td>0,725</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

In order to statistically test whether, there was a difference in mean scores from beginning to end of the day physical and cognitive stress symptoms a paired samples statistical analysis was conducted. Results show that physical and cognitive symptoms of stress increased significantly from the beginning until the end of the day ratings. Again, these results are similar to the literature in the area, suggesting that firefighter’s experience an increased in symptoms of physical and cognitive stress by the end of the day [3], probably due to stress experienced over events during the day. These findings also support the accuracy of the questions used, to assess physical and cognitive symptoms of stress among firefighters.

Finally, correlations between mean score for symptoms of physical and cognitive stress questions, shows strong positive correlations with the stress appraisal of events, suggesting that a firefighter’s experience of stress symptoms at the end of the day is strongly associated with an increased stress appraisal of events experienced during the day. Thus, the associations found between these measures, are not only in line with previous findings in the area [6], but also gives appropriateness of measures used to access the stress concept [5].

4 CONCLUSION

Overall, the Vital Analysis framework was well accepted and highly used by the firefighters in their daily routines. Results show that the questions used were reliable, and accurate enough to assess physical, cognitive symptoms of stress, and stress appraisal over events. This confirms the success of the proposed framework’s ability to support the annotation of physiological signals with the stress levels of the user. We also believe, that these findings culminate previous research limitations observed in the past when accessing stress, and the observed successful measurement of the stress concept found in this study, encourages researchers to integrate a combination of physiologic and psychological measures when investigating the concept of stress among first responders under real world conditions. Thus, findings are likely to impact future research in this area, and can also be used to design not only more efficient practical interventions, but also larger epidemiological studies.

References

An Automatic Graph-based Method for Retinal Blood Vessel Classification

Behdad Dashtbozorg\textsuperscript{1,2}  
Behdad.dashtbozorg@fe.up.pt  
Ana Maria Mendonça\textsuperscript{1,2}  
amendon@fe.up.pt  
Aurélio Campilho\textsuperscript{1,2}  
campilho@fe.up.pt  
\textsuperscript{1} INEB - Instituto de Engenharia Biomédica,  
Rua do Campo Alegre, 823, 4150-180 Porto, Portugal  
\textsuperscript{2} Faculdade de Engenharia, Universidade do Porto,  
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal.

Abstract

In this paper, we present an automatic approach to classify retinal vessels into artery and vein classes by analyzing the extracted graph from the vasculature tree and deciding on the type of intersection points (bifurcation, crossing or meeting points). The results obtained by the proposed method were compared with manual classification on 40 images of the INSPIRE-AVR dataset.

1 Introduction

Retinal vessels can be affected by many diseases. In conditions such as diabetic retinopathy, the blood vessels often show abnormalities at early stages \cite{1}. Retinal vessel dilatation is a well-known phenomenon in diabetes and significant dilatation and elongation of arterioles, venules, and their macular branches occur in the development of diabetic macular edema that can be linked to hydrostatic pressure changes \cite{2}. Changes in retinal blood vessels are also associated with hypertension and other cardio-vascular conditions \cite{3}. A sign that has been shown to be related to cardiovascular diseases is the generalized arteriolar narrowing, usually expressed by the Arteriolar-to-Venular diameter Ratio (AVR). Small changes in the AVR are associated with increases in the risk for stroke, cerebral atrophy, cognitive decline, myocardial infract \cite{4} and also it can be affected by other diseases, like diabetic retinopathy and retinopathy of prematurity \cite{5}.

In order to develop an automatic system for measuring the AVR, besides detecting the optic disc (OD) and segmenting the retinal vasculature, one of the main challenges is to classify the vessels as artery or vein (A/V classification) \cite{6}. Several works on vessel classification have been proposed in the literature \cite{7}-\cite{10}. However the automated classification of the retinal vessels into arteries and veins has received limited attention and it is still an open task in the retinal image processing field. Within this context, in this paper we propose an automatic method to classify the retinal vessels as artery or vein (A/V classification).

2 Methodology

There are some visual and geometrical features that enable the discrimination between veins and arteries. The arteries are bright red and veins are darker. In general, artery calibers are smaller than veins calibers. The arteries have thicker walls, which reflect the light as a shiny central reflex strip \cite{11}. There are methods that explore these properties for classification purposes \cite{8}-\cite{10}. Another characteristic of the retinal vessel tree is that at least in the region near the optic disc, veins rarely cross veins and arteries rarely cross arteries, but both types can bifurcate to narrower vessels and also veins and arteries can cross each other \cite{7}, \cite{11}. In our method, we assume where each intersection point is either a point where a vessel bifurcates to narrower parts or a point that a vein and an artery crosses each other.

Figure 1 depicts the block diagram of our approach for A/V classification. After vessel segmentation and centerline extraction, the main phases are: 1) graph generation; 2) graph analysis; and 3) vessel classification. The algorithm first classifies the vessels and arteries by analyzing the vessel tree as a graph which has been obtained by making a decision on the type of each intersection point (graph nodes); then, vessel segments (graph links) are classified into two different classes, and finally the A/V classes are assigned to the graph links by extracting a set of features and using a linear classifier. In the following we detail each phase of the method.

2.1 Graph generation

A graph is a simple representation of the vascular tree, where each node denotes an intersection point in the vascular tree and each link corresponds to a vessel segment between two intersection points.

We used the segmentation method proposed by Mendonça et al. \cite{12} which has good performance also in the detection and segmentation of thin vessels. An illustrative result for vessel segmentation is shown in Figure 2(b). Afterward we used the segmented image to obtain the vessel centerlines. The graph nodes are extracted by finding the intersection points (pixels with more than two neighbors), the endpoints (pixels with just one neighbor) and the high curvature points. Each link in the graph represents a connection between two nodes. The extracted graph may include some misrepresentations of the vascular structure as result of the segmentation and the centerline extraction processes. The typical errors are (1) splitting of one node into two nodes; (2) missing link on a side of a node; (3) incorrect detected link. In order to improve the accuracy, the extracted graph should be modified when one of these errors is identified. The final graph after applying the necessary modifications is shown in Figure 2(c).

2.2 Graph analysis

In the graph analysis phase a decision on the type of the nodes is made and based on node type, the links in the separate sub-graphs will be classified into one of two classes ($C_1$ and $C_2$). At the end of this phase we know which links are in the same class and in the next phase the artery/vein classes will be finally assigned to $C_1$ and $C_2$. We have considered four different types of nodes: 1) Connecting points; 2) Crossing points; 3) Bifurcation points; and 4) Meeting points.

The node classification algorithm uses the following node information: the number of links connected to each node (node degree); the direction of the links; the angles between the links; the vessel calibers related to each link and the degree of adjacent nodes.

The link classification is done for each separate sub-graph and distinct classes are assigned to each region. This means that classes $C_1$, $C_2$ will be assigned for the links in sub-graph 1, classes $C_3$, $C_4$ for the links in sub-graph 2 and so on. For each separate sub-graph, the farthest link from OD center is detected, and a class is assigned to this link (for instance $C_3$); the node connected to this link is found and based on the node type the other links are labeled as class 1 or class 2. This procedure is repeated for all nodes until there is no more unclassified links. We will repeat the process for other separate sub-graphs each time defining two new classes, until the entire graph is classified. In the end we have a classified graph with different pair of classes for each disjoint sub-graph.
3 Evaluation and results
For validating the proposed method we have used the INSPIRE-AVR dataset which contains 40 high resolution color images with 2392x2048 pixels [13]. The manual artery/vein labeling is done by an expert for all 40 images. For evaluating the classification method, we have calculated the accuracy both for centerline pixel classification and vessel pixel classification. Table 2 shows the accuracy values for centerline pixels and for vessel pixels in the entire image and also for the pixels inside the region of interest (ROI) which is the standard ring area within 0.5 to 1.0 disc diameter from the OD margin. Each row contains the accuracy values calculated using different ranges for vessel calibers. First row is the result for all the vessels, while the remaining rows present the results for the vessels with caliber higher than 5, 10, 15 and 20 pixels.

Table 2: Accuracy rates of correctly classified pixels.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Features</th>
<th>Centerline pixels in entire image</th>
<th>All vessel pixels in entire image</th>
<th>Centerline pixels inside ROI</th>
<th>All vessel pixels inside ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>Normalized Red, Green and Blue intensities under the centerline pixel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td>Normalized Hue, Saturation and Intensity under the centerline pixel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-9</td>
<td>Normalized mean of Red, Green and Blue intensities across the vessel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-12</td>
<td>Normalized mean of Hue, Saturation and Intensity across the vessel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-15</td>
<td>Standard deviation of Red, Green and Blue intensities across the vessel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-18</td>
<td>Standard deviation of Hue, Saturation and Intensity across the vessel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19-22</td>
<td>Maximum and minimum of Red and Green intensities across the vessel.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23-30</td>
<td>Intensity under the centerline pixel in a Gaussian blurred (σ = 2.4, 8, 16) of Red and Green plane.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Conclusion
We have developed an automatic method to classify the retinal vessels into arteries and veins. Accuracy values of 88.0% and 90.3% were obtained for vessel pixels of entire vascular tree and inside the ROI, respectively. The results show an improvement when compare to previous techniques and demonstrate that the proposed automatic methodology for A/V classification is reliable for AVR calculation.

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 and the research grant SFRH/BD/73570/2010.

References
Color Image Processing as a Monitoring Tool in Gas Combustion Systems

Teodoro P. Trindade
teodoro@deq.isel.ipl.pt
Edgar C. Fernandes
ecfernandes@ist.utl.pt
João M. Sanches
jmrs@ist.utl.pt

1 Instituto Superior de Engenharia de Lisboa (ISEL)
Polytechnic Institute of Lisbon (IPL)
Lisbon, PT
2 Instituto Superior Técnico (IST)
Technical University of Lisbon (UTL)
Lisbon, PT

Abstract

The applicability of image data information to estimate the spatial distribution of equivalence ratio on combustion systems is presented and discussed. The objective is to devise a numerical method capable of converting a single RGB image data, obtained by conventional CCD cameras, into a reliable sensing on local combustion state for practical atmospheric flames of CH₄ and C₃H₈ premixed gases. Flame image information need to be processed and filtered in order to extract representative data whose physical interpretation is usually not straightforward, depending on the transfer function of the overall system. To overcome this issue, is necessary for subsequent analysis an experimental reference database of flame images, obtained at controlled combustion and camera conditions. The sensing signal descriptors used in numerical method must exhibit adequate sensitivity to changes in combustion state, otherwise it may entail some information loss. Selecting multiple signal descriptors enlarges image data validation but also increases the processing time. The results obtained using a wide range of practical methane/propane flame conditions are thought to support the applicability of the proposed method for flame monitoring.

1 Introduction

The radiation emitted by flames of hydrocarbon-based fuels exhibits discrete bands in the visible electromagnetic spectrum [1]. These bands correspond to the spontaneous light emissions of excited intermedite radicals, formed along the kinetic combustion mechanism, when returning to a lower energy level, such as their ground state. This phenomena, known as chemiluminescence, is closely related to the combustion state and can be interpreted as a "signature" of a particular burning condition defined by a value of equivalence ratio φ.

\[ \phi = \frac{(Fuel - to - Air \ Ratio)_{actual}}{(Fuel - to - Air \ Ratio)_{stoichiometric}} \]  

(1)

A value of φ=1 corresponds to stoichiometric conditions, at which the reactants molar proportion of fuel to air are well balanced in such that theoretically all combustible species could be completely burned with no oxygen remaining in the products. Values of φ>1 represents fuel-rich conditions (excess fuel) and φ<1 fuel-lean conditions (excess air).

In the visible region, the population distribution of radicals CH⁺ (band head at 430 nm) and C₂⁺ (Swan system centered around 473 and 515 nm) are the main responsible for the blue-green colouration of premixed flames. Herein the symbol * indicates a combined electronic/vibrational/rotational excited level above the ground energy state of the correspondant chemical specie. Varying combustion conditions of premixed reactants, the response of the kinetic oxidation mechanism alters the radical distribution producing changes in the CH⁺ and C₂⁺ chemiluminescence intensities and thus in the colouration exhibit by flame.

Using spectroscopic techniques, this effect has been explored to monitor and control a large number of relevant flame parameters such as temperature distribution [2], location of reaction fronts [3] or the magnitude of local heat release rates [4]. The practical relevance of these light related parameters forced the development of novel monitoring techniques [5], in particular involving little demanding instrumentation requirements. In principle, flame images collected using conventional CCD (charge coupled device) cameras covers the data needed requirements in a significant range of combustion conditions [6]. As the RGB photo-sensors responds to visible wavelength and also part of the near-infrared, a digital flame image color is therefore a combined product correspondent to a broadband radiation of local emissions. Hydrocarbon flame images combines chemiluminescence radiation mainly due to CH⁺, C₂⁺ and CO₂ emitters plus black-body emission from soot particles [7]. The presence of soot on flames has a strong and markedly different dependency on equivalence ratio, degree of premixness and fuel type. In general, emissions of gas flames with φ<1.4 are dominated by black-body radiation with small contributions of chemiluminescence, which restricts the method applicability. Another limitation arise at lean flame conditions where visible chemiluminescent emissions disappear remaining only the contribution of CO₂ broad-band, whose intensity is lowered by the decline of flame temperature.

In the overall process, identification of image relevant and usable data information are thought to be the main challenges. These image features do not have an intrinsic physical meaning because the overall system transfer function in unknown, serving only as indicator of a particular combustion state. Therefore, for practical purpose any attempt to exploit a flame image data involves a previous calibration task, relating the values of the descriptor parameters with the known flame conditions.

2 Experimental Setup / Methods

A Bunsen-type burner with an outlet diameter of 20 mm was used to generate premixed laminar methane/air and propane/air flames, to be used as calibration references. Uniform gas velocity profiles of the combustion flow at the nozzle outlet were ensure by an higher ratio of contraction areas that equals 25. Fuel and air flows to the burner were metered during the experiments by precision mass flow controllers. Experimental combustion conditions using CH₄ and C₃H₈ were tested at flame power of 0.75 and 1.50 kW, in a range of equivalence ratios from 0.80 to 1.40. The imaging system used in the study was a Reflex Nikon D80 digital CCD color camera (24 bit, 3872x2592 pixels).

![Figure 1: Stoichiometric methane and propane flames: calibration profiles of G B/R descriptor values on average gray-scale intensity, Eq. (2).](image)

3 Results

It is well known that the color exhibit by an image is, among several other parameters, a function of the camera shutter speed. Variations in flame power produces the same effect of shutter speed, as they alter intensity and thus the signal distribution between red (R), green (G) and blue (B) image channels. A first step on image post-processing is the evaluation of pixel intensity (gray-scale value, I) by forming a weighted sum of the R, G and B components.

\[ I = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B \]  

(2)
The coefficients on Eq. (2) represent the average human perception of colors, being more sensitive to green and least sensitive to blue. In order to evaluate the intensity/color relations, reference experimental flame images at different combustion states were obtained at several camera shutter speeds ranging from 1 s to 1/1000 s. Using premixed methane/air and propane/air Bunsen-type flames, at fuel power of 0.75 and 1.50 kW, it was build a reference image database (0.80 ≤ φ ≤ 1.40). The output image data acquired are transferred and processed on a personal computer using a MATLAB code. Selecting a small region of interest on the reference images, covering the flame front area, the average value of intensity (gray-scale I) and signals descriptors D are determined. Being the descriptors defined by arbitrary mathematical combinations of R and/or G and/or B values. This procedure enables the production of a reference frame relating the descriptor value with the average intensity and the flame state condition φ, which only depends on fuel type (Figure 1). Several descriptors, made by arrangements of R, G and B levels, were evaluated by its characteristic profiles against I, in order to establish his adequacy on φ estimate.

Having a full descriptor distribution among gray-scale I and flame φ (Figure 2a), is then possible to estimate locally the equivalence ratio of an unknown flame by computing the pixel data in their RGB color image.

![Figure 2](image)

Figure 2: Contour maps of descriptor $D = B/G$ for premixed CH$_4$/air flames: (a) distribution of descriptor data; (b) descriptor rate of change ($\delta D/\delta \phi$).

Dark images (low intensity values) and saturated images (higher intensities) are of little interest as the descriptor values exhibits small differences among combustion states. In general, gray-scale intensities lower than 50 and higher than 225 have a limited practical use in this detection method. For a given descriptor, the best intensity range on φ detection depends on their local derivative, i.e. their rate of change with respect to φ (Figure 2b). A given descriptor D has characteristic regions of interest defining areas having larger values of $\delta D/\delta \phi$, correspondent to higher sensitivities on φ detection. In a flame image, the pixel intensities are not constant having regions of poor detection while others have a reasonable sensitivity. To overcome this problem, the equivalence ratio estimation method should apply over the same image several descriptors having complementary sensitivity regions.

The numerical procedure was tested under lean and rich (Figure 3a) conditions on domestic boiler burners. Their estimation results produces a distribution of φ along flame front that can contribute to increase the overall combustion efficiency through improvements in the burner design (Figure 3d).

![Figure 3](image)

Figure 3: Methane/air flame of a domestic boiler burner ($N_{Reynolds} = 40$): (a) flame RGB image; (b) magnification of a selected region across flame front; (c) probability function of φ on selected region; (d) estimate of flame equivalence ratio spatial distribution.

4 Conclusions

The work presented explore ways to characterize premixed gas flames using embedded information on digital flame images of methane/air and propane/air, obtained by conventional CCD cameras, as a monitoring tool for practical applications. The possibility of using RGB color model as a combustion sensor relay on the observation that CH* and C$_2$ chemiluminescence emission are linked particularly to the average values of the B and G color image channels, at rates dependent on flame fuel type.

An image database of CH$_4$ and C$_2$H$_4$ flames was experimentally obtained under controlled conditions. Using a MATLAB post-processing method and a reference numeric calibration model it was determined a set of parameters to describe the dependency between the descriptor value, the gray-scale intensity and the flame equivalence ratio, for both fuel type tested. The application of multiple descriptors to a flame image data enables the local detection of the equivalence ratio distribution along the flame front, in a range of φ between 0.8 and 1.4.

References

Abstract

Autonomous robots are becoming an integrated part of our daily life. The use of a robot for substituting the man power in different activities that might be too dangerous, repetitive or too time consuming, has become a common procedure nowadays. From autonomous vacuum cleaners, to autonomous robotic platforms used in industry, autonomous robotic space explorers and even autonomous robotic companions, robots are being developed with the purpose of imitating and improving some of the most basic human capabilities. Imitating human capabilities implies, in most applications, the implementation of a digital alternative for most of the sixth human senses. The implementation of an artificial visual sense for a robot is a research challenge that has not been yet overcome. In robotic applications that require a vision system, it is common that the artificial visual sense of the robot be the most important sense, on which all the other capabilities of the robot are based. A significant number of robotic vision systems base their functioning logic on the color information of the surrounding world. An important step in the process of “teaching” a robot the meaning of colors, is the choice of a proper color space, that could ease this task. This paper intends to be a study on the most common color spaces used in robotic applications and presents some preliminary results of an application developed with the purpose of finding the most appropriate color space to be used when implementing robotic vision systems.

1 Introduction

In the last years, autonomous robotics has been a research field under continuous evolution and expansion. Autonomous robots are being built with the purpose of easing the life of the humans, either by taking over some of the most difficult and repetitive chores of their daily life activity, or by assisting the ones with special needs. Probably the most important sense that a robot should possess in order to be able to perform its tasks in an uncontrolled environment and in an unmannned manner, is the visual sense, by means of which it should perceive the surrounding world. The human brain can process all the visual information provided by the eyes in a short amount of time since it possesses 10^{10} neurons, out of which, some have over 10000 synapses with other neurons [3]. Looking at each neuron as a microprocessor and considering that these microprocessors are able to work in parallel, the human CPU cannot even be compared to any computer that has been invented so far. Thus, providing a visual sense similar to the human one, to a robot, is yet a far to be accomplished task. Because of this, most of the robotic platforms that are being developed nowadays and that need to process visual information about the surrounding world, perform in environments that are controlled up to a certain extent, depending on the practical application of the robot.

In many industrial applications, as well as in research scopes, the processing of the visual information of a robot has as a first step, a color segmentation procedure. Especially in controlled environments, color can be an important clue for the detection of an object of interest. The color segmentation procedures imply the definition of color ranges for all the colors of interest of the application. Defining color ranges can be done by a human user, as an offline procedure, prior to the performance of the robot or it can be done online, by using automatic algorithms. The representation of the colors in digital format depends on the color space chosen. In this paper we present the preliminary results of a study on the use of color spaces in robotic vision, in order to understand if there is a more appropriate one to be used in these kind of applications. A tool for defining color ranges, both by a user and by an automatic algorithm of region growing, under different color spaces has been developed and some preliminary results have been obtained so far.

This paper is structured in five sections, the first of them being this introduction. Section 2 provides an overview on the color spaces that have been included in the testing platform. In Section 3 some preliminary results and their discussion are presented. Finally, Section 4 concludes the paper and outlines the future directions of this application.

2 Color Spaces

For the purpose of this study, four different color spaces will be studied. A color space is a mathematical model for representing the notion of color in the digital world. The conversions between these four color spaces are based on linear mathematical equations. Each of the color space has emerged at some moment in the history due to necessity of rendering images on different devices or with different infrastructures. The study that the authors are proposing, aims at finding the most appropriate color space for robotic applications.

The RGB color space 1 is the most convenient one to work with in computer graphics since it is the closest to the way the human eye works. A RGB color space is an additive color space, defined by the three chromaticities of the red, green, and blue. The main purpose of the RGB color model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography [1].

Before the electronic age, the RGB color model already had a solid theory behind it, based in human perception of colors. To form a color with RGB, three colored light beams (one red, one green, and one blue) must be superimposed (for example by emission from a black screen, or by reflection from a white screen). Each of the three beams is called a component of that color, and each of them can have an arbitrary intensity, from fully off to fully on, in the mixture. The RGB color model is additive in the sense that the three light beams are added together, and their light spectra add, wavelength for wavelength, to make the final color’s spectrum.

Figure 1: On the left, the RGB cube and on the right, an example of an additive color mixing: adding red to green yields yellow, adding all three primary colors together yields white [4].

The HSV color space is a related representation of points in an RGB color space, which attempts to describe perceptual color relationships more accurately than RGB [1, 2]. HSV stands for hue, saturation, value and it describes colors as points in a cone whose central axis ranges from black at the bottom to white at the top (Fig. 2) with neutral colors between them, where angle around the axis corresponds to “hue", distance from the axis corresponds to “saturation", and distance along the axis corresponds to “value”, “lightness", or “brightness”. The hue represents the percentage of color blend, the saturation the strength of the color and the value is the brilliance or brightness of the color.

The HSV color space is mathematically cylindrical, but it can be thought of conceptually as an inverted cone of colors (with a black point at the bottom, and fully-saturated colors around a circle at the top). Because
HSV is a simple transformation of device-dependent RGB, the color defined by (h, s, v) triplet depends on the particular color of red, green, and blue "primaries" used. Each unique RGB device therefore has an unique HSV space to accompany it.

The HSL color space is similar to the HSV one, the definition of hue and saturation, being the same as for the HSV color space. The “value” component is replaced by “lightness” and the main difference is the fact that the value, or the brightness of a pure color is considered to be the brightness of white, whereas the lightness of a pure color is the lightness of medium gray. The geometrical representation of the HSL color space is a double cone or double hexcone.

In the YUV color space, the color is represented in terms of a luminance component (Y stands for luma) and two chrominance, or color, components (U and V). This color space appeared as a necessity of introducing color television using a black and white infrastructure and encodes a color image also taking the human perception into consideration, that is, separating the luminance information by the color information.

The correctness is calculated by direct comparison with a ground truth classified image, previously created by an experienced user.

At the end of the trial, the subjects were also asked to fill in a questionnaire that would help the authors understand if there is any preferred or easier to use color space. The results show so far that the users performed faster and achieved better results in the HSV and HSL color spaces, while the RGB and YUV ones were more difficult to handle. The gathering of the results is still on-going since a large number of subjects is needed in order to have a strong conclusion at the end of the process.

The tool has been developed for the study of color spaces when performing manual color classification, as well as for studying the same color spaces when using semi-automatic color segmentation algorithms, or what the user call as supervised color classification. For the supervised color classification, several region growing algorithms have been implemented and the user only has to select a starting pixel (or a seed point) for each of the color that he wants to be classified.

For the manual classification task, the users were given an image containing different colored objects and their task was to manually define, with the help of sliders, color ranges for all the colors in the image and for all the mentioned color spaces. They were asked to take notes about their performance time, as well as about the correctness of their classification.

The results show so far that the users performed faster and achieved better results in the HSV and HSL color spaces, while the RGB and YUV ones were more difficult to handle. The gathering of the results is still on-going since a large number of subjects is needed in order to have a strong conclusion at the end of the process.

This paper presented the preliminary results of a study on the importance of color spaces in a robotic vision system. Three major issues that this study tries to address are the influence of a color space on the processing time of a color segmentation algorithm, the amount of time spent by a human user for classifying colors under different color spaces, as well as number of pixels correctly classified both manually or automatically, under the same color spaces. The collection of results for multiple human users is still in process, therefore the most important future direction is gathering these results and reaching a final conclusion about the influence of the color spaces.

This work was developed in the Institute of Electronic and Telematic Engineering of University of Aveiro and was partially funded 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 project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011).

This work was developed in the Institute of Electronic and Telematic Engineering of University of Aveiro and was partially funded 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 project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011).

References

Abstract

EEG signals have attracted the interest of scientific community for understanding how brain processes emotions. In order to extract objective conclusions, automated methods that are able to reinforce the subjective visual explorations of the signals are desirable. In this work, a feature extraction method is proposed for analysing how brain reacts to visual high/low valence stimuli and how the linked brain processes change when a novel or familiar stimulus is presented. For such purpose, experiments were carried out using the international affective picture system (IAPS) images. Global field power (GFP) from the recorded EEG signals is computed, and a support vector machine – recursive feature elimination (SVM-RFE) method is applied to the input signals. The combination of these techniques yielded up to 100% peak accuracy in both classification tasks, outperforming traditional statistical methods for group comparisons such as t-test.

1 Introduction

The dimensional model of emotions asserts that emotions can be mainly defined by two dimensions: arousal and affective valence. Some studies concentrated on one of the space dimensions for identifying the arousal intensity (high versus low) or the valence (low/negative versus high/positive), and eventually a third class neutral state. Normally, emotions are elicited by (i) presenting an external stimulus (picture, sound, word or videos) related to different emotions at some predefined interval, or by (ii) simply asking subjects to imagine different kinds of emotions.

Affective valence is an interesting, essential dimension in studies about emotion processing, which is influenced by several variables beyond the subject’s anatomy and psychophysiology. Habituation is one of these variables, and consists of a reduction in the response to a stimulus when it is repeatedly presented [1]. The interaction between habituation and affective valence processing is affected in diverse pathological and psychiatric disorders, like phobias or schizophrenia. In this work, the influence of habituation on the affective valence processing evoked by pleasant and unpleasant visual stimuli is analyzed using EEG signals.

Extracting the most relevant features linked to a specific emotional process from EEG signals is still a challenge. Amplitude and latency from the significant peaks from event related potentials (ERPs) in time domain have been usually reported in the literature [2]. However, the huge variability among subjects and the high number of trials and channels needed for the analysis often become an important drawback. In this work, we propose Global Field Power (GFP) as a representative signal of the electrical activity of the brain at a determined time instant.

Traditionally, in psychological paradigms, statistical methods have been used for comparing different populations or pairs of experimental conditions. In this study, support vector machine-recursive feature elimination (SVM-RFE), which consists of a wrapped method of feature selection [4] based on the powerful SVM classification [5] is applied. SVM-RFE has been used in very different contexts like biology or marketing [6], as well as for medical purposes using EEG data [7]. In this work, it is proposed for selecting features from the GFP signals computed from the ERPs. SVM-RFE allows to take a set of features as a whole and to work with a large number of them without requiring any statistical correction. Moreover, this algorithm for feature selection takes into account the influence of individual differences by discarding irrelevant or extreme data.

2 Material and methods

2.1 Data acquisition and pre-processing

A total of 26 female volunteers participated in the study. A number of 21 channels of EEG, positioned according to the 10–20 system, and 2 EOG channels (vertical and horizontal) were sampled at 1kHz and stored. The signals were recorded while the volunteers were viewing pictures selected from the International Affective Picture System (IAPS) picture repository, which is freely available and widely used in psychological experiments. It is composed of pictures classified by a large number of participants in terms of arousal and valence. A total of 24 high-arousal images, 12 of them corresponding to an arousal score >6, with positive valence (v = 7.29±0.65) and 12 with negative valence (v = 1.47±0.24) were selected. Each image was presented three times in a pseudo-random order and each trial lasted 3500ms: during the first 750ms, a fixation cross was presented, then one of the images was shown during 500ms, and finally a black screen followed for a period of 2250ms. The signals were preprocessed (filtered, eye-movement corrected, baseline compensated and segmented into epochs) using the NeuroScan software package. The single-trial signal length is 950ms, including 150ms previous to the stimulus onset.

2.2 Feature extraction

The high variability among subjects and even among epochs obtained from the same subject make difficult the interpretation of single trials. Averaging trials obtained from the same condition has been widely used for reducing the variability and computing the ERPs. This procedure would yield one averaged trial for each recorded channel and experimental condition. An alternative to gather the brain electrical signal occurring at a given time instant taken from the 21 EEG channels is the computation of the GFP. This measure corresponds to the spatial standard deviation, and it quantifies the amount of activity at each time point in the field considering the data from all recording electrodes simultaneously, resulting in a reference-independent descriptor for the potential field. The GFP is defined as

\[
\text{GFP}(t_i) = \frac{\sum_{j=1}^{n} \sum_{k=1}^{v} |u_i(t_j) - u_j(t_i)|}{2n}
\]

where \(u_i(t_j)\) and \(u_j(t_i)\) are the EEG signals at each electrode at the time \(t_j\) taken in all possible pairs, measured relative to a common reference, and \(n\) is the number of electrode positions used.

The GFP is computed from the ERPs. After that, for each subject GFP signals are averaged according to the conditions for making up 4 groups to be used later in classification tasks:

- Novel negative condition (NNC): Average over the GFP signals computed from EEG signals induced by negative images appearing for the first time.
- Novel positive condition (NPC): Average over the GFP signals computed from EEG signals induced by positive images appearing for the first time.
- Familiar negative condition (FNC): Average over the GFP signals computed from EEG signals induced by negative images appearing for the third time.
- Familiar positive condition (FPC): Average over the GFP signals computed from EEG signals induced by positive images appearing for the third time.

These 4 groups will allow to define four classification tasks: distinguishing brain reaction to positive/negative valence image visualization when images contain a novelty component (first time they appear) or turn to be familiar to the subject (third time the same images appear). Figure 1 shows the grand average (averaged over all subjects) GFP signals for each condition.

M.M. López
A.R. Hidalgo-Muñoz
A.M. Tomé
A.T. Pereira
I.M. Santos

IEETA, University of Aveiro (Portugal)
Dept. Experimental Psychology, University of Seville (Spain)
DETI/IEETA, University of Aveiro (Portugal)
Dept. Ciências de Educação, University of Aveiro (Portugal)
Obtained using this selection criterion are shown in Table 1. SVM classifier using again the LOO strategy. Peak accuracy values associated feature elimination loop was also included so that classification was

determined to distinguish between novel and familiar images each classification task. The first two curves show the ability of the

Figure 1: Grand average of GFP signals used for classification. The baseline points were omitted. There are four conditions in the experiment: First time a negative/positive valence image appears (NNC/NPC, respectively) and third time a negative/positive image appears (FNC/FPC, respectively).

On the other hand, modern techniques such as wrapped methods consist of including classification techniques for determining the main features to discriminate between two conditions. SVMs separate a given set of binary labeled training data with a hyperplane that is maximally distant from the two classes (or and ) known as the maximal margin hyper-plane. The objective is to build a function : [ ] 1 using training data, that is, M-dimensional patterns and class labels so that f will correctly classify new examples ( , ). Linear discriminant functions define such decision hyper-surfaces or hyperplane in a multidimensional feature space, that is:

\[ g(x) = w^T x + b \]  

(2)

where w is known as the weight vector and b as the threshold. The weight vector w is orthogonal to the decision hyperplane and b determines the distance of the plan to the origin. In [4] it was suggested a feature selection technique (SVM-RFE) that eliminates recursively the features considered less relevant by ordering the absolute values of the entries of w. This process is repeated as long as the classifier performance improves. Note that after m iterations the vector w has M M m elements as well as the feature vector x. In this work, the data vector x corresponds to the GFP signals without considering the baseline time points, that is, the initial dimension of each x is 800. After the elimination of each set of t features, the classifier is evaluated using the leave-one-out (LOO) cross validation strategy that consists of using all the samples in the dataset for training the system except one, which is used as test. This procedure is repeated N times, being N the number of samples in the dataset, after which an average value of accuracy is computed.

3 Results

Classification results obtained after applying SVM-RFE to the GFP signals are shown in Fig. 2. Each curve represents the accuracy values obtained in each iteration (m =160, t =5) of the SVM-RFE algorithm for each classification task. The first two curves show the ability of the designed system to distinguish between novel and familiar images (habituation evaluation), for each valence sign, positive and negative, respectively. The third and fourth curves represent the separability of the GFP patterns according to the sign of the valence, for familiar and novel pictures, respectively. In general, slightly higher values of accuracy are obtained for valence separation tasks than for habituation.

In order to compare the capability of the system with other traditional methods, we also applied t-test to the same sets of GFP signals as a feature selection strategy. The procedure was applied in the same conditions as the previous experiment and now only those points statistically significant (p <0.05) where chosen for classification. A feature elimination loop was also included so that classification was tested on a decreasing number of features, this time ranked by their associated p value. The selected features were then used to train/test a SVM classifier using again the LOO strategy. Peak accuracy values obtained using this selection criterion are shown in Table 1.

4 Conclusions

In this work, support vector machine-recursive feature elimination (SVM-RFE) is applied on GFP signals for selecting features in order to understand how affective valence is processed by the brain and to study the linked habituation phenomenon. GFP is a measure that considers a general spatial activity, which is interesting for studying brain emotional processes from a holistic point of view. On the other hand, the use of GFP as input space mitigates the well-known small sample size problem in pattern recognition, reducing the dimension of the input space and making it more comparable to the sample size. Our approach outperformed t-test selection when a SVM classifier is used for identifying the relevant features in the affective valence processing elicited by visual stimuli, yielding up to 100% peak accuracy for the designed classification tasks.

References


Classification of Gastroenterology images using Invariant Gabor Texture features

Farhan Riaz
http://www.dcc.fc.up.pt/~farhan.riaz/

Instituto de Telecomunicações- Porto, Porto, Portugal

Mario Dinis Ribeiro, MD

Instituto Portugues de Oncologia, Porto, Portugal

Ricardo Miguel Tavares Coimbra
Francisco Baldaque, MD

Instituto de Telecomunicações- Porto, Portugal

Karlinska Universitetssjukhus, Sweden

Miguel Tavares Coimbra

Faculdade de Medicina da Universidade do Porto
Porto, Portugal

Abstract

In this paper, we present a novel image descriptor which has been used for the classification of images from two different gastroenterology imaging modalities (chromoendoscopy and narrow-band imaging) into three different groups: normal, precancerous and cancerous. The salient feature of the proposed descriptor is its invariance to rotation, scale and illumination changes in the images. Classification results show that they produce consistent results for both imaging modalities, hinting at their possible generic use for the classification of in-body images from various imaging modalities.

1 Introduction

Gastric cancer is the second most lethal form of cancer with only a small percentage of cases diagnosed at a curable and treatable stage. Gastroenterology imaging today is an essential tool for detecting cancer effectively. It is a rapidly evolving technological area with novel imaging modalities having complementary visual features due to their different optical characteristics. Owing to the difficulties in the manual diagnosis systems, Computer Assisted Decision (CAD) systems are increasingly desirable to detect Gastrointestinal (GI) cancer effectively. In this paper, we have focused on classification of images from two imaging modalities, Chromoendoscopy (CH) and Narrowband Imaging (NBI) which have complementary visual characteristics. The former uses the full visible spectrum producing images which have richer color characteristics (Fig. 1a) whereas the latter uses two small bands of the visible spectrum (blue and green) producing images having a much poorer color resolution but richer texture features due to the enhanced visibility of its vascular patterns (Fig. 1b). From a computer vision perspective, these images pose novel challenges to the research community such as reduced color spaces, absence of geometric structures etc. for which conventional classification methods prove to be inefficient.

In this paper, we have focused on the extraction of texture features for classifying our GI images. Various state-of-the-art texture feature extraction methods [1] have been used for solving the problem of classification of images in computer vision. Given the lack of low level control on the camera, two important requirements of the image descriptors for these imaging scenarios is 1). Rotation invariance, 2). Scale invariance. In this paper, we focus on this issues and aim to propose novel features which are invariant to these variations in the images. We focus on exploiting the potential of Gabor filters for classification of our datasets, because of their advantages in joint frequency-space analysis of signals, and their physiological similarities with mammalian visual cortex. The clinical support of our work is provided by the Dinis-Ribeiro classification proposal for gastric mucosa in CH images [2] and Singh's classification proposal for Barrett's esophagus [3], which underline the features which are clinically relevant while doing diagnosis.

2 Methods

2.1 Feature extraction

We use Gabor filter for feature extraction from the images. Our choice of Gabor filters is mainly motivated by two major factors: Their capability of achieve optimal uncertainty in both space and frequency, and their similarity with primary visual cortex of mammals. A two dimensional Gabor function \( g(x, y) \) and its fourier transform can be written as [4]:

\[
g(x, y) = \frac{1}{2\pi \sigma_s \sigma_r} \exp \left( -\frac{1}{2} \left( \frac{x^2}{\sigma_s^2} + \frac{y^2}{\sigma_r^2} \right) \right) + 2\pi j W x \tag{1}
\]

\[
G(u, v) = \exp \left\{ -\frac{1}{2} \left( \frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right\} \tag{2}
\]

Where \( \sigma_u = 1/2\pi \sigma_s \) and \( \sigma_v = 1/2\pi \sigma_r \). Equation 1 is the product of a Gaussian function with complex sinusoid. This forms a bandpass filter in the frequency domain, where the bandwidth and center frequency of the filter are controlled by the standard deviation of the Gaussian function and the frequency of complex sinusoid respectively. A Gabor filter is composed of a number of self-similar wavelets. If \( g(x, y) \) is the mother wavelet, a self-similar wavelet dictionary is obtained by appropriate dilations and translations of \( g(x, y) \). An input image, \( I(x, y) \) when filtered by the set of Gabor wavelets is given as:

\[
W_{mn}(x, y) = \int I(x, y) g^*_n (x - x_1, y - y_1) dx_1 dy_1 \tag{3}
\]

Where \( W_{mn}(x, y) \) corresponds to the filter response using the filter having m\( ^{th}\) scale and n\( ^{th}\) orientation.

2.2 Autocorrelation Gabor Features

An important characteristic of Gabor filters is that if an image is rotated by a certain factor, its filter responses are represented by filters that have been rotated and scaled by the same factors. This property of Gabor filters was exploited by Kamarainen [7] who proposed the representation of Gabor responses in the form of a matrix in which the responses of rotated filters are arranged along the rows and responses of scaled filters are arranged along the columns, known as Simple Gabor Feature Space (SGFS). Consequently, for rotated and scaled images, we get matrices which are shifted representation of one another. Autocorrelation is a shift invariant operator and thus normalizes the effect of shifts in a signal. We propose the 2D autocorrelation of SGFS to obtain a novel set
of image features. Consequently, we obtain image features which are rotation, scale and homogeneous illumination invariant.

2.3 Classification

Representation of textures using filter banks is at the same time very versatile and overly redundant because images have slowly varying spatial characteristics and we do not expect filter responses to be completely different from one another at different locations. Therefore, we should get several distinct filter response vectors while all others can be seen as compositions of their noisy variations. This intuition leads us to inspect the clustering of filter responses to get a number of prototype response vectors, known as textons [5]. Our classification algorithm is divided into two stages as done by Varma [5]. Learning of texton dictionary and classification of novel images. In the texton learning stage, we select training images from our dataset and convolve them with the filter bank to generate filter responses. We use K-means clustering [6] to obtain a specific number of textons. All textons obtained from training images are collected into a single dictionary. In the testing stage, the filter response at each pixel in the novel image is quantized to the closest texton using Euclidean distance. Afterwards, the histograms showing texton occurrence frequencies corresponding to the novel images are generated. Finally, we classify these texton histograms using Support Vector Machines (polynomial kernel, one vs. one classification, sequential optimization). The objective of the classification task is to classify each image into three (CH) or four (NBI) possible categories: normal, pre-cancer and cancer. We have used 10-fold cross validation ensuring that the same image is not used for both training and testing.

3 Materials

3.1 Chromoendoscopy (CH) Dataset

The CH images were obtained using an Olympus GIF-H180 endoscope at the Portuguese Institute of Oncology (IPO) Porto, Portugal during routine clinical work. Optical characteristics of this endoscope include 140° field of view and four way angulation (210° up, 90° down and 100° right/left). The endoscopic videos were recorded during real endoscopic examinations.

3.2 Narrow-Band Imaging Dataset

The NBI images were captured using an Olympus GIF-Q160Z endoscope at Karolinska Universitetssjukhuset, Sweden during routine clinical work. Its optical characteristics include 140° field of view and four way angulation (210° up, 90° down and 100° right/left). A total of 142 chromoendoscopy images and 224 NBI images were selected from hours of video by pre-selecting images that were annotated during the real procedure by the clinician performing the exam, and later selecting each image individually for this study by an expert clinician in order to avoid intra-patient repetition and badly captured images. All images were saved as graphics files of type PNG (Portable Network Graphics).

4 Experimental Results

While learning textons, we set the number of clusters obtained using K-means clustering to 10. Therefore for CH images we obtain a total of 120 textons whereas for NBI images we obtain 160 textons (40 textons per class, 3 classes for CH and 4 classes for NBI). Consequently, we have 120- and 160-dimensional histograms as feature vectors for each CH and NBI image respectively. The Weka data mining tool (http://www.cs.waikato.ac.nz/ml/weka/) was used in the classification experiments presented here. All results were obtained using 10-fold cross validation.

4.1 Chromoendoscopy Images

For CH images, we achieved an overall classification of 82.3%; the confusion matrix is shown in Table I for a detailed analysis. An immediate observation is the system’s ability to correctly classify the images belonging to Group I and II (True positive rates of 0.88 and 0.92) however results from Group III are not promising which we suspect is because of inadequate number of images from this group (13.6% of whole dataset) for training the classifier.

<table>
<thead>
<tr>
<th>Group</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>35</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Group II</td>
<td>5</td>
<td>69</td>
<td>1</td>
</tr>
<tr>
<td>Group III</td>
<td>5</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

Table I - Confusion matrix for classification of CH images

4.2 Narrow-band images

For NBI images, we achieved an overall classification rate of 88.5%; the confusion matrix is shown in Table II for a detailed analysis. Although the overall classification for NBI images is nearly the same as that of CH images, we get the highest true positive (0.93) for NBI images belonging to Group IV (cancerous images). This shows a strong capability of the system to perform very well for classification of images belonging to cancer patients. The lowest true positive rate (0.5) is obtained for images belonging to Group III which is a vital observation again however it is pertinent to mention that these images comprise only about 5% of our whole NBI dataset.

<table>
<thead>
<tr>
<th>Group</th>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>45</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Group II</td>
<td>4</td>
<td>64</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Group III</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Group IV</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>69</td>
</tr>
</tbody>
</table>

Table II - Confusion matrix for classification of NBI images

5 Discussion

In this paper, we used novel set of textons for classification of Gastroenterology images from two distinct endoscopic imaging modalities: chromoendoscopy and narrow-band imaging. Perfect segmentation of images was assumed for our experiments. The classification accuracy results show that we obtain good performance using our novel features in a texton framework. An important observation is the consistency in the performance of these features as compared to its other counterparts though there is still a lot of room for improvement in overall classification results.

Acknowledgements

This work was supported by FCT (Portuguese Science and Technology Agency) grants SFRH/BD/45066/2008 and PTDC/EIA-CCC/109982/2009. We would like to thank FCT for its vital support for carrying out this research.

6 References

Exploring EMD for Lung Crackle Detection

Maria Beatriz Ferreira
mbeatriz@ua.pt

Guilherme Campos
guilherme.campos@ua.pt

Abstract
This paper presents a two-stage crackle detection algorithm combining Empirical Mode Decomposition (EMD) and a simple energy peak detector. A discussion is presented of the main issues arising in the implementation of the EMD stage and the solutions adopted. The algorithm was implemented in MATLAB® and preliminarily tested on an annotated 10-second respiratory sound file, without any prior systematic training. Applying the energy peak detection to the intrinsic mode function of order 3 (IMF3) generated by the EMD stage, an 87% F performance index was achieved.

1 Introduction
The analysis of pulmonary sounds is an important means of diagnosing respiratory pathologies [1]. The aim is to detect artefacts, typically referred to as adventitious lung sounds (ALS), regarded as potential symptoms of lung disease. Signal processing and computing technologies can contribute to improve the diagnosis techniques based on stethoscopy, reducing ALS detection subjectivity.

This paper addresses the automatic detection of crackles [2], an important category of ALS, related to chronic fibrosis, pneumonia and chronic bronchitis [3]. Crackles are short-duration (<20ms) nonstationary sounds, with a frequency range normally between 100 and 2000 Hz. The energy ratio of crackles to normal respiratory sound is low, resulting in significant waveform distortion. The wide variation in magnitude, explosive nature and broad spectrum observed in crackle waveforms and the fact that they can overlap makes for an intricate signal processing challenge.

Numerous detection methods have been proposed in the literature, applying a variety of algorithms (time-domain analysis, wavelet transforms, fractal dimension filtering, fuzzy logic, neural networks…) and often combining them in more or less complex ways [1]. However, with systematic validation still largely unaddressed, there are serious concerns about their practical applicability. The challenge of reliable automatic crackle detection remains open.

Huang et al. presented Empirical Mode Decomposition (EMD) as a new technique to analyse nonstationary and nonlinear signals [4]. This naturally suggested the possibility of applying it to crackle detection, an idea explored by Charleston-Villalobos et al. [5] and Hadjileontiadis [6].

In contrast to other methods as discussed in [4], EMD offers the advantages of being adaptive, local, intuitive and a posteriori.

The core of EMD is an iterative decomposition process, known as sifting, whereby data is reduced into so-called Intrinsic Mode Function (IMF) components based on local properties.

An EMD-based crackle detection algorithm will comprise two stages:
1) Generation of the IMF components, which maintain the signal’s energy distribution and thus the physical meaning of the data;
2) Filtering of the chosen IMFs to find crackle positions.

In the work of Charleston-Villalobos et al. [5], visual inspection of the generated components (particularly IMFs 2 and 3) indicated a good match between the most prominent features and the corresponding signal annotation data. Although encouraging, these results were only qualitative, as stage 2 was not addressed i.e. no automatic method of detecting and counting crackles was implemented. The paper also acknowledged difficulties in telling crackles from basic respiratory sounds and the lack of reliable validation data.

Hadjileontiadis [6] completed the algorithm by combining EMD and Fractal Dimension (FD) filtering. The FD filter stage analyses the temporal evolution of the generated IMF waveform complexity, and this information is used as a pointer to crackle positions.

This paper explores an alternative path to overcoming the limitations of [5], avoiding the hybrid approach of [6], which seems overly complex. Instead, stage 2 is reduced to a simple energy peak detector (based on the premise that the most prominent peaks match crackle positions) and the focus is firmly placed on solving EMD implementation issues in order to fully expose the potential of the technique itself.

2 The EMD algorithm
EMD is an adaptive process that empirically separates the data into IMFs, applying a time-varying filter. Instantaneous frequency and energy are the relevant global variables of EMD [4].

Each IMF contains an intrinsic oscillatory mode and the instantaneous frequencies can be defined anywhere in this function. The process of signal decomposition into IMFs is called sifting. To form an IMF, a time series must satisfy two conditions:
1) Considering the whole data set, the number of local extrema must be equal to the number of zero crossings or differ from it by one;
2) At any point, the mean of the upper and lower envelopes, defined by the local maxima and minima, respectively, must be zero.

The sifting process is necessary because “misbehaved” signals (such as crackles) may contain multiple instantaneous frequencies at a time. It involves subtracting the higher oscillation modes and iterating on the residual, as illustrated in the flowchart of Figure 1, until signal decomposition is complete.

![Figure 1: The Sifting process](image)

The end result is a representation of the original signal in terms of a set of N IMFs and a residue:

\[ s(t) = \sum_{n=0}^{N} IMF_n(t) + r_N(t) \]

Equation 1: Expansion of a signal \( s(t) \) in terms of IMFs

It should be stressed that the physical properties of a time series are maintained when it is expanded into its IMF components; the decomposition is complete, orthogonal, local and adaptive [4].

3 EMD-based crackle detection
The detection algorithm was developed in MATLAB®. It takes a respiratory sound file as its input and generates an annotation file, recording the detected crackle endpoints sequentially.

Stage 1 implements the sifting process (see Fig. 1), generating IMFs up to the order specified by the user. First, the local extrema (maxima and minima) and zeros of the input signal are obtained. The upper and lower envelopes of the signal are then obtained through interpolation and used to compute the mean. Subtracting this mean from the original signal yields the first IMF candidate, \( h_1(t) \); if this new signal verifies the two conditions required (section 2), it forms an IMF and the sifting process continues, taking the residue as a new input signal. Otherwise, the process is repeated taking \( h_1(t) \) as the input. This recursive procedure
terminates when the required number of IMF components and the respective residue are obtained, forming a complete expansion of the original signal.

The following stage is a straightforward energy peak detector whose operation is illustrated in Fig. 2. It starts by squaring the IMF signal and applying a smoothing convolution filter. The position of the highest peak of the resulting signal is considered the midpoint of a crackle and used to split the signal into two segments. The procedure is then applied to each of them and repeated recursively in order to detect lower amplitude peaks down to a pre-defined energy threshold.

![Figure 2: Excerpt of respiratory sound (a) with corresponding IMF3 (b) and crackles indicated by detected IMF3 energy peaks (c).](image)

### 4 EMD implementation issues

Important issues mentioned (but not explained in detail) in EMD-related literature became apparent in the implementation of this algorithm. One of them was the end effect in the spline fitting, extremely notorious in the initial attempts to obtain IMFs. It causes wide variation at the edges of the signal, distorting the decomposition. Boundary conditions were implemented to attenuate this problem: initially, the first and last data points are considered both as minima and maxima, and the remaining extrema of the set are computed. Subsequently, the current values of those first and last minimal/maximal are updated if a result derived from the slope defined by the following/previous points is less/higher than the current value.[7]

Another issue is related to the tolerance and threshold parameters of the sifting algorithm, which need to be adapted (fine-tuned) to the data set under analysis. In this case, the tolerance (parameter tol in Fig. 1) was empirically set, based on the maximum absolute value of the signal. Regarding the stopping criteria, the mean criterion proposed by Rilling, G. et al.[8] was applied. This technique, based on two thresholds for an amplitude near zero, considers small global fluctuations and at the same time large local excursions, preventing over-iteration.

Several tests led to the conclusion that running both IMF test conditions in the respiratory sound signals took too long and the results obtained by verifying only one of them were identical to those obtained when both were tested. Hence, in order to avoid over-sifting and excessive execution times, the algorithm only observes the mean criterion, which seems to be enough for these specific data sets.

The final issue is the choice of the IMF to be used as input to the peak detector. Visually comparing each IMF with the reference sound data annotated by health professionals, IMF 3 and IMF 4 seemed to provide the best match. Lower and higher order IMFs appeared to contain mostly high frequency and basic respiratory sound noise, respectively. This only partially agrees with the observations of Charleston-Villalobos et al. [5].

### 5 Results

Preliminary performance tests were carried out on a 10-second respiratory sound file annotated by a health professional, without prior training to optimise the algorithm’s parameters. By adjusting only the algorithm’s energy peak detection threshold, an F index as high as 92% (harmonic mean of SE=94% and PPV=89%) was achieved.

![Figure 3: Graph showing the relation between the detection performance and the energy threshold](image)

### 6 Conclusions and Future Work

This paper explored the application of EMD in automatic respiratory crackle detection systems. Outstanding EMD implementation issues and alternatives to solve them were analysed in this specific application context. A fully functional EMD-based automatic detection algorithm was developed. Highly promising results were achieved in pilot validation tests, calling for further, more systematic performance evaluation on more extensive test sets. Moreover, numerous algorithm refinement possibilities can be envisaged, including parameter optimisation through training, pre-filtering and combination of different order IMFs and/or IMF derivatives. Since they are associated to different oscillation modes, different order IMFs might also be helpful in fine/coarse crackle classification.

### References

Abstract

Given the large number of dynamic geometry systems (DGS), geometry automated theorem provers (GATP) and repositories of geometric information (constructions and/or conjectures), we face the need of a query mechanism for formal descriptions of geometric constructions. The DGS and GATP describe the geometric constructions using formal languages where the elements and the relations between them are described formally and not in terms of a given geometric model. Given a formal language we need to be able to look for similar construction, sub-constructions or even different construction sharing a common property, e.g., a set of constructions about right angled triangles. Our approach is to transform the geometric construction into a semantic graph representation of the construction, in a given ontology. Graph pattern recognition algorithms can then be used to search for the similarities we need and the results brought back to the geometric setting.

1 Introduction

Dynamic geometry systems (DGS) [11] distinguish themselves from drawing programs in two major ways, the first is their knowledge of geometry. Indeed, from a initial set of objects drawn freely in the Cartesian plane (or maybe, on another model of geometry), one can specify/construct a given geometric figure using relations between the objects, e.g., the intersection of two non-parallel lines, a line perpendicular to a given line and containing a given point, etc. That is, one uses a DGS by constructing a geometric figure with geometric objects and geometric relations between them, not placing points on specific cartesian coordinates. Another major feature of a DGS is its capability to introduce dynamics to a given geometric construction. Given the fact that one specifies/constructs a geometric figure using a set of basic elements, e.g. points, lines, circles and relations between them, the DGS allows its user to move one of the basic (free) elements form its initial placement to another placement in the Cartesian plane, the relations will be kept, so a movement in a single point can entail the movement of almost all the other elements in the construction, i.e., when moving a basic object, we will move that object and all the other elements that are related to it, always preserving the geometric properties of the construction.

Most (if not all) DGS possess a formal language for the specification of geometric constructions. In some systems this formal language is explicit, in others it is hidden from the user by the graphical interface. The intergeo project designed a common format, called t2G, for this formal language which is already accepted by many DGSs [1,9].

Geometry automated theorem provers (GATP), being formal systems, need a formal language to describe geometric conjectures. GATPs are nowadays mature tools capable of proving hundreds of geometric conjectures [3]. There are two major lines of research: the synthetic proof style and the algebraic proof style. The algebraic proof style begins by reducing the geometric properties to algebraic properties expressed in terms of Cartesian coordinates, proving the theorem by pure algebraic methods, so they do not belong to the realm of the geometric reasoning. The synthetic proof style GATP uses geometric reasoning, and its formal language is an extension of the formal language used by the DGS. The t2GATP project goal is to define a common language, an extension of the t2G language, to the DGS/GATP tools [7].

The design of common languages, and the emergence of Web repositories of geometric knowledge is an attempt to make widely available the already vast data set of geometric knowledge. The intergeo project [5], the GeoThms [8] and the TGTP [6] systems already meet some of these goals having provided a large data set of geometric information widely available. In these systems the question of querying the geometric-construction is not solved, that is, it is not possible to query the data set for a construction similar to some other construction, or to query for all constructions having some common geometric properties.

The goal of our research is to develop a search mechanism for geometric constructions (done by a DGS or a GATP) using the formal specification of the construction. Our approach is to transform the geometric construction into a semantic graph representation of its elements and relations, in a given ontology. The resulting semantic graphs can then be used, using graph pattern recognition algorithms, to detect matching constructions, and the results can be presented to the user once converted back to their geometric representation.

1.1 State of the Art

The term “Geometric Pattern Matching” refers to the recognition of shapes in a given set of points and it is an important area of research with applications in computer image processing, manufacturing, robotics, VLSI design, military intelligence, etc. with many articles written under many different approaches, but that it is not the goal we pursue. By geometry construction recognition we mean the search for pattern not on a given model of geometry but in the formal specification of the construction, or, equivalently, in its semantic graph representation, for a given ontology. As far as we known this is an area still to be explore, we are making the first steps in that direction.

2 Geometric Constructions and their Semantic Graph Representations

The DGS and most of the GATP are based in a constructive geometry [10]. A construction is specified stating an initial set of points, the free points, implicitly universally quantified over a field of characteristic different from two, and from these initial elements, using a set of constructive rules the geometric construction is built [2,4]. For rendering the DGS will attach to the free points some Cartesian coordinates, but these coordinates will be ignored by the GATP.

Let us consider a (very) small example (see Fig. 1).

Points A, B and C are the free objects of this construction, starting with them we define new elements A′, B′ and C′ as being the middles of the opposite line segments, and P as the intersection of AA′, BB′, CC′ and hence the barycenter of the triangle. The GeoThms and the TGTP systems share a database with more then 180 constructions like this one, specified in a constructive geometry language, the geometric queries are a much needed feature for those systems.
To represent this construction we need to chose an ontology. Most ontologies will contain the concept of "point," but, depending on the theorem to prove, one may add concepts for "line," "line segment," "angle," "length," etc. Typical relations are "belongs to," "is parallel to," etc. In our case, since points $A'$, $B'$ and $C'$ are defined as being middles of line segments, we need the relation "is middle of;" which is a superrelation of "belongs to," in the partial hierarchical order of the ontology.

In Fig. 2 one can see the semantic graph of construction 1 in the ontology using points (represented by □) and line segments (repar. by △) as concepts, and relations "belongs to" (→) and "is middle of" (dashed →).

Figure 2: Semantic graph of construction 1. (Labels have been placed only to allow the reader to establish a connection with the original geometric figure — they are not part of the semantic graph per se.)

Fig. 2 represents construction 1 entirely, in the given ontology. We could also represent the same construction in an ontology with relation "has same length" instead of "is middle of;" in that case we would use "has same length" arrows between nodes $AC'$ and $BC'$ in both directions, instead of the "is middle of" arrow between $C'$ and $AB$.

Changing ontologies will cause a transformation of semantic graphs, and this is an operation that can be done mechanically.

3 Querying Geometric Constructions

Every user query can be transformed in a pair consisting of a (minimal) ontology and a (small) semantic graph. For example, the query find a construction containing a triangle would result into the ontology ["point" (□), "line segment" (△), "belongs to" (→)] and the semantic graph

$$G = \begin{array}{c}
\quad \\
\quad \\
\quad \\
\end{array}$$

By representing all constructions in this ontology (which, in the case of Fig. 2, would require to replace the relation "is middle of" by the weaker relation "belongs to"), we only need to seek for the presence of $G$ as a semantic subgraph of the representation of each construction.

Here is another example: to find all constructions containing an intersection of three line segments (or three lines, in a slightly different ontology), one would need to seek the presence of the following subgraph:

$$G' = \begin{array}{c}
\quad \\
\quad \\
\quad \\
\end{array}$$

4 Conclusions and Future Work

Having already defined a semantic graph representation counterpart for the GC geometric constructions and a way of extracting information from it we need to define a query language that can be used in the Intergeo, GeoThms and TGTP (and others) systems, i.e., we need a user interface enabling users to make queries about geometric constructions and/or properties. The support for the geometric languages (i2G and i2GATP) and the trimming of the search mechanism to be used on the large set of geometric construction is also important.

After a query has been submitted to the engine, it will need to find the (minimal) ontology needed, and, if necessary, convert constructions into it. Then it will detect semantic subgraphs and either display them to the user, or display the corresponding parts of the geometric constructions, or simply reply with a boolean value for constructions matching the query.

An interesting research direction to pursue in the future, is also to investigate how inference used in proofs by GATPs, interacts with the semantic representation.

References

Compression of DNA microarrays using a mixture of finite-context models

Luis M. O. Matos
luismatos@ua.pt
António J. R. Neves
an@ua.pt
Armando J. Pinho
ap@ua.pt

Signal Processing Laboratory
IEETA/DETI
University of Aveiro,
3810-193 Aveiro, Portugal

Abstract

DNA microarray images are experiments that allow the identification of the function and regulation of a large number of genes in a single experiment. These microarray experiments consist of a pair (green and red channel) of 16 bits per pixel grayscale images. This paper addressed a lossless method to compress these kind of images, using a mixture of finite-context models and arithmetic coding. We use a 3D context configuration, expectation-based bitplane coding and also a typical image context template in the mixture. We conclude that using a mixture of finite-context models we are able to improve the compression results.

1 Introduction

We can use DNA microarray technology to study and monitoring gene function across a large number of genes, and even entire genomes. The raw data resulting from a microarray experiment consist of a pair of 16 bits per pixel grayscale images (see Figure 1). These images usually require tens of megabytes to be stored or transmitted without any compression. Due to this fact, and the need for long-term storage and efficient transmission, a lossless compression method with progressive decoding capabilities is an important challenge. In the literature we can find several proposals for lossless compression of microarray images [1, 4, 5, 6, 7]. In [3] we can find a review of DNA microarray image compression, where the authors describe the most relevant approaches published in the literature.

(a) Green channel (b) Red channel

Figure 1: Example of a pair of images (1041 × 1044 pixels) that results from a microarray experiment.

2 Method

The proposed method is based on a mixture of finite-context models, where several models are used to estimate the probability of each symbol. Each model generates a probability estimate and the final probability is a weighted sum of the probabilities estimated by all models used.

2.1 Finite-context models

A finite-context model assigns probabilities to the symbols of an alphabet \( \mathcal{A} \), according to a conditioning context. The probability estimates \( P(X_{n+1} = s | c^i) \) are calculated using symbol counts that are accumulated while each bitplane of the microarray image is compressed. We use the estimator

\[
P(X_{n+1} = s | c^i) = \frac{C(s | c^i) + \alpha}{\sum_{a \in \mathcal{A}} C(a | c^i) + |\mathcal{A}| \alpha},
\]

Parameter \( \alpha \) allows balancing between the maximum likelihood estimator and an uniform distribution. For \( \alpha = 1 \), (1) is the well-known Laplace estimator. The per symbol information content average provided by the finite-context model of order-\( k \), after having processed \( n \) symbols, is given by

\[
H_{k,n} = \frac{1}{n} \sum_{i=0}^{n-1} \log_2 P(X_{i+1} = s | c^i) \text{ bps},
\]

where “bps” stands for bits per symbol.

2.2 Expectation-based bitplane Coding (EBC)

In 2010, Chen et al. [2] proposed a lossless compression algorithm that uses EBC. EBC is a strategy that is useful in some bitplanes where the pixel values are not so meaningful as they seem. Sometimes, the neighboring bits are only refining bits that fine-tune the value of a particular pixel. This model can be used in our approach in microarray images. Suppose that \( p = \{p_1, p_2, \ldots, p_k\} \), where \( p_i \) is the value of the pixel at bitplane \( i \). When we are encoding the \( n \)th bitplane, if the pixel \( p \) is already encoded at the current bitplane, its expectation value is formulated as

\[
E(x) = \sum_{i=0}^{15} 2^i \left(2^{n-1} - 1\right).
\]

On the other hand, if the pixel \( p \) has not been encoded at current bitplane, its expectation value is defined as:

\[
E(x) = \sum_{i=0}^{15} 2^i \left(2^n - 1\right).
\]

During the compression process, we use this expectation values instead of the real bit values. In case of using the EBC, if the expected value of the context is lower than the expected value of the current pixel, the context bit used is 0, otherwise 1 is used. The variable size template used is described in Figure 2 and as we can see, there are two future pixels (pixels 4 and 3) that are used in the context template. The pixel denoted by 0 represents the current pixel being compressed. Using this approach, we can select future pixels for the context template, which is more efficient for the least significant bitplanes, where the neighboring bits of each bitplane are merely refining bits and they only fine-tune the final value of that particular pixel.

![Figure 2: Context template used in EBC.](image)

2.3 Proposed approach

As already said, the proposed method is based in a mixture of finite-context models. The algorithm processes the microarray images bitplane by bitplane using several models. We used the 3D finite-context model (see Figure 3) proposed in [6], a typical context template (see Figure 4) and also the EBC model presented in the previous subsection. The goal of our approach is to compute a probability estimate of several models.
and to combine them into a single probability that is used to compress each symbol. Each model contributes to the final probability estimate of the next outcome symbol. Therefore, we can compute the probability estimate using a weighted average of the probabilities provided by each model, according to

\[ P(x_{n+1}) = \sum_k P(X_{n+1} = s' | c^k) w_{k,n}, \]

where \( w_{k,n} \) denotes the weight assigned to model \( k \) and

\[ \sum_k w_{k,n} = 1. \]

Figure 3: 3D context configuration proposed in [6].

We use a typical context template configuration similar to the template presented in Figure 4. The proposed algorithm supports several template configurations at the same time.

![Context Template Configuration](image)

Figure 4: An example of context template configuration used in our approach.

### 3 Results and Conclusion

In order to be able to compare our results with the results presented in [6], the microarray images used were collected from three publicly available sources. 1) 32 images that we refer to as the APO.AI set; 2) 14 images forming the ISREC set; 3) three images previously used in MicroZip. The image sizes range from 1000 × 1000 to 5496 × 1956 pixels and all of them have 16 bits per pixel. The average results presented take into account the different sizes of the images, i.e., they correspond to the total number of bits divided by the total number of image pixels.

In Table 1, we have the compression results of 3 standards (JPEG2000, JPEG-LS and JBIG), the algorithm proposed in [6] and our approach (rows “Mix10” and “MixEBC”). The “Mix10” corresponds to a mixture between the 3D context configuration presented in Neves [6] and the context template configuration presented in Figure 4. As we can see, there are small improvements in all datasets for the “Mix10”. On the other hand, the “MixEBC” which correspond to a mixture between the same 3D context configuration of “Mix10” and EBC, generates worst compression results. Table 2 presents another interesting metric, which is the number of bits per pixel that would be required if we could select always the best model to encode each symbol (but without considering the side information indicating the model used).

Table 1: Compression results in bits per pixel for the proposed method (“Mix10” and “MixEBC”), and the method presented in [6]. We included also the results for 3 image compression standards (JPEG2000, JPEG-LS and JBIG).

<table>
<thead>
<tr>
<th>Method</th>
<th>APO.AI</th>
<th>ISREC</th>
<th>Microzip</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>JBIG</td>
<td>11.851</td>
<td>10.925</td>
<td>9.297</td>
<td>10.393</td>
</tr>
<tr>
<td>JPEG-LS</td>
<td>10.608</td>
<td>11.145</td>
<td>8.974</td>
<td>10.218</td>
</tr>
<tr>
<td>Mix10</td>
<td>10.302</td>
<td>10.194</td>
<td>8.662</td>
<td>9.610</td>
</tr>
<tr>
<td>MixEBC</td>
<td>10.316</td>
<td>10.214</td>
<td>8.674</td>
<td>9.625</td>
</tr>
</tbody>
</table>

Table 2: Lower bound of each proposed method. These values represent the number of bits per pixel that would be required if we could select the best model to encode each pixel (but without considering the side information indicating the model used).

![Lower Bound Table](image)

That, despite using the EBC model in the mixture, it does not provide better results, although in a competitive approach it could be more efficient than the “Mix10” approach.

The proposed method, based on a mixture of finite context-models, as we can see, provides small improvements compared to the method [6]. As future work, it will be interesting to explore other models than the EBC in the mixture. Also, we could use some pre-processing techniques in the microarray images in order to improve the compression results.

### 4 Funding

Supported by the European Fund for Regional Development (FEDER) through the Operational Program Competitiveness Factors (COMPETE) and by the Portuguese Foundation for Science and Technology (FCT) in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference Pest-C/EUI/UI0127/2011).

### References


Abstract
This paper presents a comparative study of several segmentation methods applied in ultrasound (US) images. This comparative study will be performed to extract the bone contours in US images. With this study is possible to identify the drawbacks, advantages, and analyze the accuracy and performance of each method. Six different methods are implemented to segment bones, in images of real human femurs.

1 Introduction
Image segmentation is a fundamental aspect of image processing and computer vision. Recently, extensive studies have been made and many techniques have been developed.

The main purpose of segmentation is to divide the original image into homogeneous regions for one or more characteristics. Each of the regions has the ability to be processed separately for extracting important information of images. In medical imaging processing, the segmentation identifies anatomical location of organs or some regions of interest, where is important intervene. For example in surgical planning is important to identify with high level of accuracy, where is the localization of bones or tumors.

There are a large number of segmentation techniques that have been proposed and implemented but there is still no standard that meets all possible criteria when applied to unspecified image types. In this paper is tackled the US image, that is intrinsically noisy.

Most effective segmentation algorithms are obtained by careful combinations of parameters. These algorithms should be adjusted to the image and to the characteristics of the anatomical structure to be segmented.

Segmentation is beneficial when applied to image data of both patients with pathology and normal volunteers. Scans of people without pathological abnormalities can be used as a method for comparison to define abnormality.

This paper is tackles bone segmentation in US images and is organized as follows. In section 2, the Methods of segmentation adopted in this work are presented. The next section describes the Experimental results. Finally, conclusions and future work are presented.

2 Methods
Ultrasound image acquisition was performed using a portable echograph with a 5 MHz linear probe, coupled to a computer, using a frame grabber. The images (10 slices) are representative of the femur bone of an adult person, showing different parts of it. These images have a high level of noise, i.e., speckle, depicted in Figure 1.

After acquisition, is important to define the region of interest (ROI), depicted in Figure 1. For these images is of special interest to identify where is the bone to perform the segmentation. Figure below represent a square around the region of interest, defined in the initialization step.

At this stage is necessary, in accordance with the acquired image, to find segmentation algorithms capable to obtain reliable results in US images. Six algorithms will be compared, in order to find which one presents less error as compared with the femur bone contours extracted manually by the medical expert. For the femur bone contours extracted the paper will answer the following question: What is the Best Segmentation Method that meets our needs?

Several state-of-the-art methods were analyzed and it was decided to use six methods, four of them implemented in the Insight Toolkit (ITK), that provides the ability to further develop each method and customize a complete application. The last two methods are implemented in Matlab.

3 Experimental Results
The Insight Toolkit (ITK) is an open-source software toolkit for performing registration and segmentation. Segmentation is the process of identifying and classifying data found in a digitally sampled representation. Typically the sampled representation is an image acquired from such medical instrumentation, US images in this paper.

Matlab is a high level programming language, is interactive and efficient for numerical calculation, where one of the basic data structures is based on the matrix representation, adequate to work with images.
For comparing the described methods a performance criterion must be used, which allows us to determine the best one for the task at hand. As a performance criterion was used the root mean squared error (RMSE) (equation 1), where \( V_g \) are the ground-truth values (obtained by the medical expert manually), \( V_e \) are the estimated values (obtained after the application of the methods), \( n \) is the number of 2D image points contained in both the contours.

\[
RMSE = \sqrt{\frac{\sum(V_g-V_e)^2}{n}} \tag{1}
\]

Since \( V_g \) and \( V_e \) can have different sizes, a cubic spline was used to interpolate both the contours, allowing having the same size \( n \).

A second performance criterion was used, e.g., the Mahalanobis distance (MAHA), equation 2. It is based on correlations between variables by which different patterns can be identified and analyzed. It gauges similarity of an unknown sample set to a known one, and differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant.

\[
MAHA = \sqrt{(V_g - \mu_g)^T S^{-1} (V_e - \mu_e)} \tag{2}
\]

where \( S \) is the covariance matrix, between \( V_g \) and \( V_e \), and \( \mu_g \) and \( \mu_e \) the respective mean values.

The methods described above were applied to a set of 10 US images (576 x 720) in various sections of the femur. After applying the six methods to each one of the 10 US images, was obtained a set of points corresponding to the upper contour of the bone for each of the images. These points are used for calculating the RMSE and MAHA, allowing a comparison of methods. For each image (set of points) was found the best method i.e., the one with lower RMSE and MAHA. These results are presented in Table 1. The TLSS (Threshold Level Set Segmentation) method achieved better results in 60% of the test images, using the RMSE metric. For the Mahalanobis distance, the TLSS method also presented the best results for 80% of the test images. These results showed that level set based methods are the best suited for bone contour segmentation in US images.

Figure 2 presents the bone extracted from the image presented in figure 1, and also image number 8 in Table 1, using the TLSS method. The contour is defined as the upper part of the extracted bone, i.e., the first white pixel seen in the vertical direction (up-down).

Figure 3 depicts the contour points \( V_g \) (red) and \( V_e \) (blue), respectively the ground-truth and estimated, showing the excellent result obtained.

Table 2 presents the results obtained to image 8, the image with the lowest RMSE presented in table 1. This result shows that the TLSS method outperforms the other methods for at least 1 pixel (the MRIS method).

<table>
<thead>
<tr>
<th>Image Number</th>
<th>RMSE</th>
<th>Method</th>
<th>MAHA</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.4154</td>
<td>TLSS</td>
<td>1.3498</td>
<td>TLSS</td>
</tr>
<tr>
<td>2</td>
<td>5.7708</td>
<td>SDS</td>
<td>1.6852</td>
<td>TLSS</td>
</tr>
<tr>
<td>3</td>
<td>4.8076</td>
<td>FMS</td>
<td>1.8066</td>
<td>GACS</td>
</tr>
<tr>
<td>4</td>
<td>7.4789</td>
<td>SDS</td>
<td>1.5439</td>
<td>GACS</td>
</tr>
<tr>
<td>5</td>
<td>2.9446</td>
<td>TLSS</td>
<td>1.6999</td>
<td>TLSS</td>
</tr>
<tr>
<td>6</td>
<td>2.7766</td>
<td>TLSS</td>
<td>1.6316</td>
<td>SDS</td>
</tr>
<tr>
<td>7</td>
<td>2.8813</td>
<td>FMS</td>
<td>1.7827</td>
<td>TLSS</td>
</tr>
<tr>
<td>8</td>
<td>1.538</td>
<td>TLSS</td>
<td>1.7100</td>
<td>TLSS</td>
</tr>
<tr>
<td>9</td>
<td>2.9180</td>
<td>TLSS</td>
<td>1.7659</td>
<td>TLSS</td>
</tr>
<tr>
<td>10</td>
<td>5.6189</td>
<td>TLSS</td>
<td>1.3856</td>
<td>TLSS</td>
</tr>
</tbody>
</table>

4 Conclusions

The paper presented a study to find the best method for bone femur extraction, based in US images extracted from an experimental setup presented in our laboratory.

For the six methods studied the TLSS, Threshold Level Set Segmentation, based in the ITK implementation outperforms the other five methods. The TLSS method also have the advantage of needing only a seed point, inside the bone, when compared to some other methods that need a set of points around the bone contour, for example the MRIS method (Minimization of Region-Scalable Fitting Energy for Image Segmentation) that also presented good results. In general, the results showed that level set based methods are best suited for bone contour segmentation in US images.

In the future it is expected to further reduce the mean square error in order to minimize the segmentation errors. These results will also be used in tracking algorithms to obtain a bone surface in real time.

References

Segmentation of the Lungs in PET Scans: A Watershed-based Approach

Fábio Nery¹
fabionery@gmail.com
José Silvestre Silva²
jsilva@ci.uc.pt
Nuno C. Ferreira¹,³
nuno@ibili.uc.pt
Francisco Caramelo²
fcaramelo@ibili.uc.pt
Ricardo Faustino¹,³

Abstract

The vast amount of data generated by state-of-the-art medical imaging techniques leads to the need for the development of automatic approaches for its analysis requiring minimal to no user interaction. The definition of the area of interest, using image segmentation algorithms, is a frequent pre-processing step in computer-aided diagnostic procedures.

This paper presents a marker-driven watershed segmentation algorithm for the segmentation of the lungs using solely physiological information, retrieved from PET scans.

The ground-truth for the assessment of the segmentation results consists in manual delineations performed by two experts. The lung areas which are the output of our algorithm are then compared to their correspondent ones, obtained by a manual delineation, using several figures of merit.

The segmentation results are promising. The algorithm-expert variability is comparable to the inter-observer variability, which is significant, given the physical limitations and lack of anatomical information provided by PET.

1 Introduction

Positron Emission Tomography (PET) is a nuclear medicine imaging technique that quantifies the biodistribution of positron emitting radiotracers in vivo. Thus, it allows the retrieval of significant information regarding the physiological properties of an organism. Although the main applications are focused on the field of Oncology, where the glucose analog \(^{18}\)FDG (2-[\(^{18}\)F]Fluoro-2-deoxy-D-Glucose) plays a pivotal role, many new radiotracers are being developed, expanding the ability of PET to study a wider array of physiological phenomena. The segmentation of PET images presents some additional challenges that are a lesser concern in other modalities, such as Computed Tomography. Many limitations arise due to the intrinsic nature of the technique such as an increased noise component, lower spatial resolution and blurring effect. Therefore, most works deal with the segmentation of tumors \([1, 4, 6, 7, 8, 9, 10]\), since their higher glucose uptake makes them easily discernible, when compared to other structures. Nevertheless, other authors addressed the segmentation of different structures, such as the liver \([5]\) and myocardium \([3]\). Furthermore, semi-automatic methods where the resulting segmentation depends on the user input, have also been developed \([2]\). As far as we know, this is the first time that a fully automatic approach based in the watershed segmentation concept is developed with the goal of extracting lung boundaries using only PET data.

2 Marker-driven Watershed Segmentation

2.1 Theory

A gray-level image can be visualized as a topographic surface. Besides the two dimensions of an image, the pixel intensities determine the “height” of the surface. This surface is then flooded from below (regional minima) and as the water rises, dams are built to prevent the merging of water coming from two distinct catchment basins. Eventually, only the tops of the dams are “visible” above the water level, corresponding to the connected boundaries extracted by the algorithm.

The major issue when applying this concept is the frequent oversegmentation, which most of the times seriously jeopardizes the results. The solution for this problem is to use segmentation markers, which allow to identify the foreground (objects of interest) and the background. Besides “controlling” the segmentation procedure, markers are an effective approach to introduce a priori knowledge to the segmentation procedure.

2.2 Algorithm Description

This section presents the main steps carried out by our proposed approach. In order to suppress some noise, the algorithm starts by applying a \(5 \times 5\) Wiener adaptive filter (Fig. 1 (a)). As we previously mentioned, the use of markers is extremely important. After the first filtering operation, the algorithm is ready to compute the different types of markers.

2.2.1 Internal Marker Definition

The first step to identify the markers corresponding to the foreground, also known as internal markers, is to apply two thresholds \((T_{\text{low-uptake}}\) and \(T_{\text{body}}\)), which respectively identify low-uptake structures (where at least parts of the lungs are included) and the border of the patients’ body. The structures identified by \(T_{\text{low-uptake}}\) are accepted or rejected according to a distance criteria, defined as the distance, \(D\) (given in pixels), between the border of the patients’ body and the closest “large” object after filling the image resulting from \(T_{\text{low-uptake}}\) where the seeds are the interior pixels adjacent to the border of the body. Morphological operations are used to smooth the obtained objects.

2.2.2 External Marker Definition

The definition of the markers is by no means complete without the identification of the structures belonging to the background, which we want to avoid. Two main problems arise from the segmentation using only the internal markers, both being caused by the poor definition of the lung boundaries. For each problem, a different type of external marker was defined. The first problem is the leaking to areas far from the lung boundaries. To solve this, the first type of external marker is derived from the previously computed border of the patients’ body (subsection 2.2.1). The set of pixels corresponding to this marker is obtained by performing \(0.25 \times D\) dilations of the body border and keeping only the inner contour of the resulting object.

The second problem is the merging of the lungs into a single object. To avoid this to happen, we use the binary image containing the internal markers, compute its distance transform (Euclidean distance) and then we obtain the watershed lines of the distance transform image. When the internal markers are composed of more than two objects, this procedure yields several lines. Therefore, we devised specific criteria to keep only the line that correctly vertically divides the image between both lungs.

Having defined all the necessary markers, we compute a smoothed version of the gradient of the filtered image (Fig. 1 (b)), to enhance the edges and then the markers are imposed into the gradient image (Fig. 1 (c)). The watershed lines of the resulting image are then computed, obtaining the lung boundaries (Fig. 1 (d)).
3 Evaluation

The results of the algorithm for 60 computed lung contours were compared to their correspondent ones, which are manual delineations by two experts, which we consider to be the gold standard of the segmentation results. Two complementary figures of merit were used to quantify the similarity between the areas enclosed by the resulting contours:

\[
\text{Sensitivity} = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{1}
\]
\[
\text{Specificity} = \frac{N_{TN}}{N_{TN} + N_{FP}}, \tag{2}
\]

where \(N_{TP}, N_{FN}, N_{TN}, N_{FP}\) are, respectively, the number of True Positive, False Negative, True Negative and False Positive pixels. The sensitivity figure of merit quantifies the amount of true positives that are correctly identified as such. On the other hand, the specificity measures the proportion of true negatives that are correctly identified as such.

4 Results and Discussion

The results of the comparisons between the algorithm results and delineations by both experts, as well as the inter-observer comparison are conveniently presented using box-plots (Fig. 2) which allow a fast perception of the overall dispersion of the data and the presence of outliers. The naming convention is the following: A→Algorithm E→Human Expert. An immediate conclusion is that the specificity is very high (>0.97), higher than the sensitivity (>0.8 excluding outliers), meaning that the algorithm performs better in the determination of true negatives than true positives. This is consistent to what we concluded by visual inspection of the segmentation results, which was a slight overestimation of the lung area by our algorithm. Most of the outliers correspond to the apex region of the lung. This may be explained by the degrading effect of the respiratory movement in this particular region. Nevertheless, the algorithm-expert variability is comparable to the inter-observer variability, indicating promising results.

Acknowledgements

The authors would like to thank ICNAS for providing the PET studies.

References

Study of cerebrovascular dynamics using functional MRI

Joana Pinto\textsuperscript{1}
João Jorge\textsuperscript{1,2}
Pedro Vilela\textsuperscript{3}
Patrícia Figueiredo\textsuperscript{1}

\textsuperscript{1}Institute for Systems and Robotics, Instituto Superior Técnico, Technical University of Lisbon, Portugal
\textsuperscript{2}Laboratory for Functional and Metabolic Imaging École Polytechnique Fédérale de Lausanne, Switzerland
\textsuperscript{3}Imaging Department, Hospital da Luz, Lisbon, Portugal

email: joanasequeirapinto@gmail.com

Abstract

The study of cerebrovascular dynamics is an important component in the evaluation of different neurological diseases. Non-invasive functional MRI techniques are currently being explored for this purpose, namely Blood Oxygenated Level-Dependent (BOLD) and Arterial Spin Labeling (ASL). Here, we aimed to characterize the spatiotemporal dynamics of both ASL and BOLD responses to a breath-holding (BH) challenge. Images were obtained from a group of 7 healthy subjects on a 3T MRI system. The ASL and BOLD data were analysed based on a general linear model using two different approaches: Sine Delay Optimization (SDO) and Sine Cosine Linear Combination (SCC). Three contrasts were obtained from these analyses: BOLD, BOLD\textsubscript{cw} and cerebral blood flow (CBF). Cerebrovascular Reactivity amplitude (CVR) and cerebrovascular delays (CVD) were estimated using both approaches, for each type of contrast. Both SDO and SCC models were successful in obtaining CVR and CVD maps. However, SCC yielded CVR and CVD on a voxelwise basis and showed relatively greater sensitivity for both BOLD and BOLD\textsubscript{ASL} responses. Although no significant improvement was obtained for CBF data, the CVD voxelwise results showed larger variations across the brain relative to CBF data. Further work is required in order to optimize the methodology for the analysis of CBF BH data and to clarify the different delay distributions observed for BOLD and CBF.

1 Introduction

Magnetic Resonance Imaging (MRI) provides a completely non-invasive technique for cerebrovascular reactivity (CVR) measurement. In particular, the Blood Oxygenated Level Dependent (BOLD) contrast is commonly used to measure this parameter as the response to vasoactive respiratory challenges \cite{1}. However, the BOLD contrast does not reflect a single physiological process, but rather represents the combined effects of several parameters \cite{2}. On the contrary, Arterial Spin Labeling (ASL) provides an appealing alternative contrast by only depending on the effect of changes in cerebral blood flow (CBF), allowing a more direct physiological parameter estimation.

Studies using breath-holding (BH) hypercapnic tasks have successfully measure CVR in a non-invasive manner \cite{3}. A correct quantification of this parameter can be helpful in the assessment of several vascular pathologies and the associate response delays can be of great interest in the characterization of vascular latencies across the brain. These temporal dynamics and spatial variations of both BOLD and CBF BH responses remain largely unexplored \cite{4,5} and only a few studies have simultaneously measured the CBF response using Arterial Spin Labeling.

Here, we aim to characterize the spatiotemporal dynamics of both the CBF and BOLD responses to a BH challenge, by estimating the corresponding CVR and associated cerebrovascular delays (CVD) using two different approaches: Sine Optimization Delay (global) and Sine Cosine Linear Combination (voxel-by-voxel).

2 Material and Methods

2.1 Image Acquisition

A group of 7 healthy subjects (4 males; 25.7±4.8 years old) was studied on a 3T Siemens Verio scanner using a 12-channel radio frequency coil. BOLD images were obtained using a gradient echo–echoplanar imaging (GE-EPI) sequence, with TR/TE=2500/25ms. For the BOLD signal, two paradigm-related regressors were defined, a sine wave and a cosine wave both with a 75s period (paradigm frequency) and with phases ranging from 0s to 75s. For the ASL protocol, a full perfusion signal GLM was defined \cite{7}, using three regressors. The first regressor modelled the baseline control-tag alternating intensity variation. The second regressor modelled the intrinsic BOLD\textsubscript{ASL} signal present and it was used a sine wave with a period of 75s and phases ranging from 10s to 30s. The final and third regressor was formed by multiplying the control-tag and the paradigm-related BOLD\textsubscript{ASL} regressors together, and modelled the activation component of the control-tag signal (CBF weighted signal).

SDO is a global approach. Parameter estimation consists on finding one optimized global CVD value for the whole brain and a associated CVR map. A general linear model (GLM) was defined for each type of data, BOLD and ASL. For the BOLD signal, only one paradigm-related regressor was defined, a sine wave with a 75s period (paradigm frequency) and with phases ranging from 0s to 75s. For the ASL protocol, a full perfusion signal GLM was defined \cite{7}, using three regressors. The first regressor modelled the baseline control-tag alternating intensity variation. The second and third regressors modelled the intrinsic BOLD\textsubscript{ASL} signal present and it was used a sine wave and one cosine wave with a period of 75s and phases ranging from 10s to 30s. The final and third regressor was formed by multiplying the control-tag and the paradigm-related BOLD\textsubscript{ASL} regressors together, and modelled the activation component of the control-tag signal (CBF weighted signal).

SCC is a voxel-by-voxel approach. The parameters CVD and CVR were estimated on a voxel-by-voxel basis. Again, a GLM was defined for each type of data. For the BOLD signal, two paradigm-related regressors were defined, a sine wave and a cosine wave both with a 75s period (paradigm frequency). For the ASL protocol, a full perfusion signal GLM was defined using five regressors. The first regressor modelled the baseline control-tag alternating intensity variation. The second and third regressors modelled the intrinsic BOLD\textsubscript{ASL} signal present and it was used one sine wave and one cosine wave with a period of 75s. The third and fourth regressors were formed by multiplying the control-tag and the paradigm-related BOLD\textsubscript{ASL} regressors (sine/cosine) together, and modelled the activation component of the control-tag signal (CBF weighted signal).

In both approaches, SDO and SCC, parameter estimates corresponding to each regressor were determined. Positive and negative contrasts of interest were defined and t-tests were then applied to the BOLD, BOLD\textsubscript{ASL} and CBF positive contrasts, yielding a Z-statistic map for each one. Additionally, for the SCC approach, F-tests were also applied to the BOLD, BOLD\textsubscript{ASL} and CBF positive contrasts. The maps were then thresholded using cluster thresholding with cluster p<0.05 and with voxel Z>2.3 (for BOLD and BOLD\textsubscript{ASL} data) and Z>0.5 (for the CBF data). GLM regression coefficients were then used to estimate the parameters CVR and CVD.

In the SDO approach, the CVR amplitudes were estimated by calculating the BOLD and CBF percent signal change relative to the corresponding baseline. A global optimized delay was selected by

![Figure 1: Breath-holding paradigm comprising eight 75 seconds cycles of 20 seconds BH and 55 seconds normal breathing for Arterial Spin Labeling (ASL) data and three 75 seconds cycles of 20 seconds BH and 55 seconds normal breathing for Blood Oxygenation Level Dependent (BOLD) data.](Image 1)
choosing the value of delay that yielded the maximum number of voxels exhibiting significant BH response. Relatively to the SCC approach, CVR values were estimated by calculating the BOLD and CBF percent signal change of sine/cosine combinations, relative to the respective baselines. CVD values were estimated by calculating the angle of the sine/cosine combinations. Contrarily to the SDO approach, SCC estimations were done on voxel-by-voxel basis.

3 Results

The results for one illustrative subject are shown in Figures 2 and 3.

BOLD, BOLD\textsubscript{ASL} and CBF responses exhibited similar globally delays across the brain (BOLD $28\pm6s$, BOLD\textsubscript{ASL} $28\pm7s$, and CBF $28\pm9s$). However, CBF delays varied much more across the brain.

The total number of responsive voxels and mean CVR values are shown in Figure 4 and 5, respectively, for the subjects group. The number of voxels exhibiting significant BH responses and the mean brain CVR values were significantly increased for BOLD and BOLD\textsubscript{ASL} data using the SCC relative to the SDO approach. CBF data showed no significant improvements.

4 Conclusions

A sine/cosine model was proposed to estimate on a voxel-by-voxel basis the amplitude and delay of both CBF and BOLD responses to a BH paradigm. This was compared with a sine model with global delay optimization. The proposed sine/cosine model was successful in improving the sensitivity of both BOLD and BOLD\textsubscript{ASL} responses to the BH paradigm. Although no significant improvements were found for CBF data, cerebrovascular delay maps were obtained for the first time in this case, showing large variations across the brain. These results are partly due to the much lower SNR of CBF data. Further work is required in order to optimize the methodology for the analysis of CBF BH data and to clarify the different delay distributions observed for BOLD and CBF.

Acknowledgements

This work was supported by FCT project PEst-OE/EEI/LA0009/2011.

References

A Total Variation based Denoising Algorithm for 3D Ultrasound

Marya V. Afonso
mfonso@isr.ist.utl.pt

João M. R. Sanches
jmrs@isr.ist.utl.pt

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisbon, Portugal

Abstract

This paper presents an algorithm for denoising a three dimensional image, under the Rayleigh distributed multiplicative noise model, which is the observational model for Ultrasound imaging. The proposed method performs a variable splitting to introduce an auxiliary variable to serve as the argument of the 3D total variation term. This leads to two problems involving the data fidelity term and the regularizer, which are solved alternately leading to an instance of the Alternating Direction Method of Multipliers (ADMM) method, for which convergence is guaranteed. This framework offers a simple way to apply the TV regularizer as at each iteration, one needs to minimize the sum of a TV term and a least squares term, as against the Rayleigh penalty function. This framework can further be extended to decompose the image and speckle.

1 Introduction

Ultrasound (US) has emerged as a popular medical imaging modality in a number of medical imaging applications because of its lower cost, wide reach, flexibility, lack of radiation, and intra-operability. In Brightness mode (B-mode) 2D US, a two-dimensional image is acquired by a linear array of transducers simultaneously scanning a plane through the body. However, 2D imaging only allows views in the same plane as the US acquisition is performed. Another limitation is that the 2D US image can manually position and orient the probe. Due to the impracticality of acquiring the 3D volume of the carotid artery to be reconstructed from a series of 2D slices. The slices can be acquired mechanically wherein the probe position is controlled by a motor which sweeps it over the volume in a predetermined manner, or freehand wherein the user can manually position and orient the probe. Due to the impracticality of slicing over the entire volume, voxel values of the 3D object are available for only a small set of the 3D positions.

1.1 Problem Formulation

We assume that the dimensions of the volume, that is, the number of voxels is $L \times M \times N$. We will represent the volume as a vector in lexicographic ordering, $x \in \mathbb{R}^{n}$, where $n = L \times M \times N$ is the number of voxels. Assuming linear mechanical scanning along the height, we denote the set of observed 2-D slices as $S_{i} \in \mathbb{R}^{M \times N}$, $i = 1, \ldots, n_{s}$. The acquired 2-D slices are of size $M \times N$, and the number of slices is $n_{s} \leq L$. Hence the number of observed voxels is $m = n_{s}MN$. Knowing the scanning pattern over the volume, we obtain the position coordinates of each pixel in the observed slices. We can therefore define an observed vector $y \in \mathbb{R}^{m}$, which contains the voxel values from the observed slices after mapping to the volume and ordering in lexicographic order. In the case of scanning along a linear line, the slices can be simply stacked. In the more general case where the acquisition is random and freehand, this sampling pattern can be applied, knowing the positions and orientations of the acquisition patterns.

Knowing the coordinates of the pixels of the 2-D slices, we can define a 3-D binary array which is at the voxels corresponding to pixels in observed slices and 0 elsewhere. This leads to a linear observation model for modeling the acquisition of slices from the volume

$$y = A x,$$

where the linear operator $A \in \{0, 1\}^{m \times n}$ multiplies the vector representing the 3-D volume with the binary mask, and discards the voxels that are not sampled. It is essentially the $n \times n$ identity matrix with $n - m$ rows removed. For a denoising problem, it is equal to the identity matrix $A = I$, which is the case when all voxel values are observed.

In US images, the noise is multiplicative and is called speckle,

$$y = (A x) \eta,$$

where $\eta$ is the speckle field that is multiplied element-wise with the observed volume. For the Radio Frequency envelope images, it is assumed that the observed image follows Rayleigh statistics, with the likelihood

$$p(y|x) = \prod_{i=1}^{m} \frac{y_{i}}{|A x|_{i}} \exp \left(\frac{-y_{i}^{2}}{2|A x|_{i}}\right),$$

where $(A x)_{i}$ is the $i^{th}$ element of the sampled vector $A x$, and $y_{i}$ is the $i^{th}$ element of the noisy observed vector $y$.

To formulate the optimization problem to estimate the volume $x$, given $y$ with the Rayleigh multiplicative model (3), we apply a logarithmic transformation,

$$f = \log(x),$$

and then use the convex data fidelity term from [8], leading to the total variation (TV) regularized convex problem

$$\min_{f} \sum_{i} \left(\frac{y_{i}^{2}}{2} e^{-f_{i}} + f_{i}\right) + \alpha TV(f),$$

where as before, $f_{i}$ represents a voxel from the vector representation of the volume $f$, indexed by the coordinates $(i_{x}, i_{y}, i_{z})$. $\alpha > 0$ is the regularization parameter, and the regularizer function is the 3D TV function which is given by

$$TV(f) = \sum_{i} \left(\frac{(f_{i,x,j,z} - f_{i-1,y,j,z})^{2} + (f_{i,j,y,z} - f_{i,j-1,z})^{2} + \cdots + (f_{i,j,z} - f_{i-1,y,z-1})^{2}}{2}\right)^{\frac{1}{2}}$$

Although there exist fast solvers for the sum of a quadratic data fidelity term and a TV term [3, 7], we cannot apply them to the multiplicative and Rayleigh noise model.

1.2 Contributions

This work proposes an algorithm to solve the denoising problem (5), based on the Augmented Lagrangian (AL)/Alternating Direction Method of Multipliers (ADMM) framework [5] which decompose the problem into a series of simpler problems, are fast, and have guaranteed convergence conditions.

2 Proposed Method

We find the solution of problem (5) by using an approach based on the AL/ADMM framework. In [2], an ADMM based method was presented for solving a problem similar to (5) but without logarithmic compression, for the case of Synthetic Aperture Radar (SAR) imaging. We perform a variable splitting [4] and introduce an auxiliary variable $u$ to serve as the argument of the TV term, with the constraint $f = u$. This leads to the constrained optimization problem,

$$\min_{f, u} \sum_{i} \left(\frac{y_{i}^{2}}{2} e^{-f_{i}} + f_{i}\right) + \alpha TV(u),$$

subject to $f = u$. (7)
Using the augmented Lagrangian, this problem can be shown to be equivalent to the minimization problem,
\[
\min_{\mathbf{u}} \sum_{i} \left( \frac{\gamma^2}{2} e^{-f_i} + f_i \right) + \alpha TV(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{f} - \mathbf{u} - \mathbf{d}\|_2^2, \tag{8}
\]
where \( \mu \geq 0 \) is called the AL penalty parameter, and \( \mathbf{d} \) is the so-called Bregman update vector \([7]\). The AL algorithm iterates between minimizing the objective function in (8) with respect to \( \mathbf{f} \) and \( \mathbf{u} \), leading to a Gauss-Seidel process (for more details, see \([1]\)) which at iteration \( k \) is summarized as,
\[
\mathbf{u}^{k+1} = \arg\min_{\mathbf{u}} \alpha TV(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{f} - \mathbf{u} - \mathbf{d}^{k}\|_2^2, \tag{9}
\]
\[
\mathbf{f}^{k+1} = \arg\min_{\mathbf{f}} \sum_{i} \left( \frac{\gamma^2}{2} e^{-f_i} + f_i \right) + \frac{\mu}{2} \|\mathbf{f} - \mathbf{u}^{k+1} - \mathbf{d}^{k}\|_2^2, \tag{10}
\]
\[
\mathbf{d}^{k+1} = \mathbf{d}^{k} + \mathbf{u}^{k+1} - \mathbf{f}^{k+1}.
\]

Problem (9) is a quadratic denoising problem and is solved using a 3D implementation of Chambolle’s algorithm \([3]\). The objective function in (10) is separable for each voxel \( f_i \), and can be decomposed into \( n \) problems,
\[
f_i^{k+1} = \arg\min_{f_i} \frac{\gamma^2}{2} e^{-f_i} + f_i + \frac{\mu}{2} (f_i - u_i^{k+1} - d_i^{k})^2, \tag{11}
\]
which can be solved efficiently using a few iterations of Newton’s method \([2]\). Even though (9) and (10) are not solved exactly, the convergence conditions for ADMM \([5]\) only require that their error sequences decrease monotonically. The final estimate of the volume is \( \hat{x} = \exp(\mathbf{f}) \).

### 3 Experimental Results

We first demonstrate the proposed method on a synthetic example, in which a synthetically generated cylinder is corrupted with speckle, and then for denoising real B-mode US images of a carotid artery, stacked to form a volume. In the synthetic example, the generated cylinder is the ground truth and the mean square error can be computed relative to it. All experiments were performed on MATLAB on an Ubuntu Linux based laptop, with the Intel i5 processor and 8 GB of RAM.

The synthetic cylinder is a volume of size 128 \( \times \) 128 \( \times \) 128 and consists of ones in the volume of the cylinder, and zeros elsewhere. A cross-section in the xy-plane is shown in Fig 1(a). After speckling, the corresponding slice from the noisy volume is shown in Fig 1(b). Fig 1(c) shows the corresponding slice after denoising using the proposed method, and Fig 1(d) shows the denoised cylinder in a 3D view. The mean square error (MSE) between the original synthetic volume \( \mathbf{x} \) and the denoised volume \( \hat{\mathbf{x}} \), was 0.003, and the CPU time taken was 55.64 seconds.

The US images of the carotid artery were acquired transversally over a region of length 8 cm. There were 60 slices of size 255 \( \times \) 256, each roughly corresponding to an area of 3.9 \( \times \) 4 sq.cm. Fig 2(a) shows a B-mode image (which has speckle), and Fig 2(b) shows the corresponding slice from the denoised volume, shown in Fig 2(c). The CPU time was 162.59 seconds.

### 4 Conclusions

We have proposed an ADMM based algorithm for denoising images with multiplicative speckle noise. Preliminary results show that the proposed method is accurate and computationally efficient. Extension for reconstruction with a sub-sampling of the slices, speckle estimation, and comparison with other solvers 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 Postdoctoral fellowship (contract no. SFRH/BPD/79011/2011) and FCT project (PEst-OE/EEI/LA0009/2011).

### References

Estimating world coordinates in perspective vision systems for humanoid robots

João Silva
joao.m.silva@ua.pt

Nuno Lau
nunolau@ua.pt

António J.R. Neves
an@ua.pt

Transverse Activity on Intelligent Robotics
IEETA/DETI University of Aveiro
3810-193 Aveiro

Abstract

The use of digital cameras on humanoid robots is very common. To extract information from camera images on these kind of platforms, we need to know the pose of the camera relative to our objects of interest. In the soccer application we project all the points of interest on the ground plane, in a 2D top view over the field of play. To accomplish this projection based on the pinhole camera model, we need to extract a set of values about the camera pose, based on the current position of the robot body and the camera position on the robot head. This document present a brief summary of the steps involved in such operation and the tool developed to support such steps.

1 Introduction

The use of cameras as main sensory information source for robots is nowadays almost a requisite, due to the flexibility of such source. To extract information from digital camera images, several methods of image processing are widely used. This results in the location of objects on the image, which are pixel coordinates in a 2D representation of the world. In this document, we present the approach we use to extract information of point location on the world from the pixel coordinates provided by the vision analysis. The scenario used is robotic humanoid soccer on the RoboCup competition. The RoboCup is an international joint project to promote robotics and artificial intelligence. It includes several leagues, several of them related to soccer. Within these leagues, the Standard Platform League (SPL) is the one where this work is applied, a league where all the teams use the same robots, the humanoid NAO. These robots are equipped with 2 cameras which, in the version we use, do not have any overlap on their fields of view and cannot work simultaneously. Moreover, the robot controller is a Geode processor with a frequency of 500Mhz and possesses 256 MB of RAM. In the SPL, the objects of interest we use on the field are projected at the ground level. It is only thanks to this “restriction” that we are able to tackle the problem with a single camera without the need for more information.

2 Analysis of pinhole camera model

To accomplish the ground point projection, we use the pinhole camera model to analyze the geometrical relations of coordinates in 3D space and their projection into the camera 2D CCD. The first step was to make a static analysis of a vision system, in a way similar to the one presented in [1]. Figure 1 presents two schematics of a camera system, side view and semi top view.

The several measurements involved are:

- \( h_{\text{offset}} \) - height of the camera relative to the ground;
- \( r_{\text{offset}} \) - radial distance from the camera to the robot center;
- \( \alpha_{\text{offset}} \) - angular offset from the vertical axis to the camera center axis;
- \( x_{\text{offset}} \) - distance from the center of the robot to the point in the center of the image, projected on the ground;
- \( \alpha_{\text{an}} \) - angle measured between \( \alpha_{\text{offset}} \) axis and \( \text{pixel}_n \);
- \( \beta_{\text{an}} \) - angle measured between the robot frontal axis and \( \text{pixel}_m \);
- \( \text{distance}_n \) - distance, measured on the ground, from the center of the robot to \( \text{pixel}_n \) projected on the ground;
- \( \text{distance}_m \) - distance from the center of the camera to \( \text{pixel}_m \) projected on the ground;
- \( \text{distance}_{\text{diag}} \) - distance between lens and CCD;
- \( \text{pixel}_n \) - number of pixels \((n)\) along a CCD column;
- \( \text{pixel}_m \) - number of pixels \((m)\) along a CCD row;

\[
\alpha_{\text{offset}} = \tan^{-1} \left( \frac{x_{\text{offset}} - r_{\text{offset}}}{h_{\text{offset}}} \right) \tag{1}
\]

From (1) we can generalize to an angle \( \alpha_{\text{an}} \):

\[
\alpha_{\text{an}} = \alpha_{\text{offset}} - \tan^{-1} \left( \frac{\text{distance}_n - r_{\text{offset}}}{h_{\text{offset}}} \right) \tag{2}
\]

Given the geometric properties of a camera vision system, we can relate an angle \( \alpha_{\text{an}} \) and a pixel along a vertical column of the CCD,

\[
\text{pixel}_n = \frac{\tan(\alpha_{\text{an}}) \times \text{focal length}}{\text{pixel height}} \tag{3}
\]

From (2) and (3), we obtain

\[
\text{pixel}_n = \tan \left[ \alpha_{\text{offset}} - \tan^{-1} \left( \frac{\text{distance}_n - r_{\text{offset}}}{h_{\text{offset}}} \right) \right] \times \frac{\text{focal length}}{\text{pixel height}} \tag{4}
\]
With some manipulation of 4, the distance corresponding to each pixel can be found by

\[
distance_{x_n} = r_{offset} + h_{offset} \times \tan \left( \alpha_{offset} - \frac{\text{pixel}_n \times \text{pixel height}}{\text{focal length}} \right)
\]

(5)

We can thus obtain the distance, from the robot center, of any point on the image, using (5).

Following a similar analysis and based on the schematic of Fig. 1, we know relation between \( \alpha_{bot} \) and distances \( distance_x \) and \( distance_y \) projected on the ground:

\[
\alpha_{bot} = \tan \left( \frac{distance_y}{distance_x - r_{offset}} \right)
\]

(6)

Since we have a height associated, and the horizontal angle has a relation with both the distances (XX and YY), we can derive the following

\[
distance_y = \text{pixel}_n \times distance_{diag} \times \frac{\text{pixel size}}{\text{focal length}} \times \cos \left( \alpha - \tan \left( \frac{distance_x - r_{offset}}{h_{offset}} \right) \right)
\]

(7)

3 NAO vision system

The initial analysis was based on a system with a fixed camera, with its vertical axis aligned with the robot front and horizontal axis parallel to the ground. In the case at hand, the NAO robot, the camera is placed on the head, which is a mobile part of the robot. Thus, in this case, we have to consider that the camera has rotation on all three axis XX, YY and ZZ, commonly known as roll, pitch and yaw respectively.

Looking at a camera in space, and considering XX the camera focal axis with the YYZZ plane on the CCD, we can see that if we apply a roll angle to the camera, we have a rotation of the pixels. The first step of the coordinates estimation is to rotate the given pixel by the roll angle of the camera to correct that distortion.

The pitch analysis basically results in the \( \alpha_{offset} \) that allows us to get the coordinates relative to the current image. Through expressions (5) and (7) we estimate the coordinates of the given pixel considering that the front of the robot is the current direction of the camera.

Finally, knowing the yaw of the camera, we can rotate the given point, already in ground plane coordinates, and obtain the final coordinates, relative to the robot front.

3.1 Extracting camera angles

To get the necessary angles of the camera on the space, we make use of the robot kinematics, combined with the use of the inertial unit present on the robot. Moreover, each camera of each robot has different angles relative to the head, since the construction of the robot does not guarantee precise values. Thus, we need to check the cameras roll, pitch and yaw angles on the head.

To help in extracting this information and to visually confirm both these parameters and the results of the estimations presented, a visual tool was created. This tool has a set of parameters necessary to calculate the ground coordinates of the image pixels and can run with a live feed from the robot camera, as well as a live feed from the robot sensors which allows us to verify also the kinematic model estimations, in an offline, static scenario. For the initial stage of the tool, we solved (5) and (7) in relation to \( \text{pixel}_x \) and \( \text{pixel}_y \) so that we can project a grid with known distances on the image pixels according to our model. Figure 2 presents some screen shots of the application with such a grid. In the example, the squares have a 36mm side and the bottom line is at 311 mm from the robot center, manually measured with tape. The robot is placed in such a way that it is parallel to the chessboard.

4 Results

The developed tool was a need that came up when starting to extract the pixel ground projection information. The main need was initially to help the development and eventual corrections of the expressions, based on the knowledge of what the grid should be like over the image of a known environment. It is also a tool to try and test the angular parameters of the camera, also allowing to test the intrinsic parameters focal length, pixel size and CCD center.

Figure 3 illustrates the projection of points over the center circle of the soccer field when applying the expressions over the theoretical camera placement on the robot head and with the correction estimated through the camera calibration tool.

References

Robotic Soccer: 
a real challenge for cooperative robotics

Luis Ferreira
lff@ua.pt
Ricardo Dias
ricardodias@ua.pt
João Cunha
joao.cunha@ua.pt
António Neves
an@ua.pt
José Luís Azevedo
jla@ua.pt
Bernardo Cunha
mbc@det.ua.pt

Transverse Activity on Intelligent Robotics
IEETA/DETI, Universidade de Aveiro
Aveiro, Portugal

Abstract
This paper gives an overview on the challenges presented in the RoboCup competitions and how them can improve cooperative robotics in a non-controlled environment. Below a solution is presented by the CAMBADA team, a robotic soccer team created for the RoboCup Middle Size League (MSL), to deal with the MSL challenges. This paper presents an overview of the internal software structure of the CAMBADA agents, mainly in what concerns coordination and cooperation among the agents, and the proposed solution for the RoboCup 2012 MSL Technical Challenge, where the CAMBADA team attained the first place.

1 Introduction
To overcome great obstacles, people have learned that cooperative work is a powerful tool, and many work nowadays have to be done in teams. However, it is not just the number of workers that dictates the success of the operation, it is also their ability to communicate with each other and coordinate their efforts. With the increasing availability of robots in different areas of human activity, it is natural that they start to mimic human behaviour, becoming more independent and capable. Progresses in wireless communication technologies enables teams of robots to share information in real time and coordinate themselves, while longer lasting batteries and smaller components make possible the creation of simple robots for domestic use.

Because technology is still not in a level where humans can rely completely on robots for daily tasks, there has to be another reason to encourage researchers to develop better algorithms and better equipments. In this area, the RoboCup competition acts as a catalyst for teams to show each year new discoveries and innovations on specific areas, such as robotic soccer, domestic challenges, rescue simulations and others. By starting with a specific goal in a controlled environment, and then increasing the complexity yearly, researchers face one challenge at a time instead of being confronted with multiple problems at once, which would increase the difficulty degree exponentially.

Keeping in mind the participation in these challenges, the University of Aveiro created the CAMBADA team (acronym of Cooperative Autonomous Mobile Robots with Advanced Distributed Architecture), a robotic soccer team that participates in the RoboCup Middle Size League (MSL).

The CAMBADA coordination model is based on sharing the world state between all robots, by the use of their communication capabilities and a real time database. The decision architecture is based on Roles, that provide a way to perform high-level tasks. Each role is a combination of Behaviours that compose the high-level tasks [1].

This paper presents the effectiveness of Roles on the CAMBADA software architecture as well as their influence in the RoboCup 2012 MSL Technical Challenge, were the team achieved first place. The organization of the paper is the following: Section II explains the software architecture responsible for the coordination of the robots and the available roles. Section III presents the MSL technical challenge for 2012, the rules and the created roles for the team to complete the challenge. Finally, Section IV discusses the results and Section V concludes with final thoughts and future work.

2 Coordination Between Robots
The CAMBADA robots follow a distributed approach, both in their hardware architecture and their software architecture [1]. The hardware is distributed in three layers which facilitate replacement and maintenance. The software is constituted by five processes executed concurrently, and these communicate by means of a Real Time Database(RTDB) [2] which is physically implemented in shared memory. An illustration of the software architecture can be seen in Fig. 1.

Being responsible for the internal communication of the five processes, the RTDB is one of the most important features of the software architecture. It contains essential state variables divided in two sections, a local section that holds the data needed by the local processes and is not to be broadcasted to other robots, and a shared section which is divided between all running agents and contains sub-sections with data of the world state as perceived by each one. Each agent transmits its own shared section, thus keeping all RTDBs with updated information. Moreover communication is done wirelessly by means of Wifi.

The choice of making the robots communicate by Wifi is the best solution because it is a technology that has proven to be robust and adaptable to many situations, also there is a large range of hardware that supports it.

Figure 1: The CAMBADA software architecture
The ability that each robots has to communicate to the rest of the team their data is an important asset when making decisions, as each robot is capable of deciding which role they are assuming, knowing their teammates roles. This makes another important feature of the CAMBADA software architecture, a dynamic role-based architecture with formation [1].

There are two main situations during a game, freeplay and set pieces. In freeplay are used 2 roles: the striker which is the robot closer to the ball and the midfield which assumes a defensive position behind the striker. In set pieces a combination of 2 roles is used, the replacer which passes the ball and the receiver who receives it. In both situations there is always another robot performing the goalie role.

Besides the dynamic roles, there is also a dynamic formation. Using Delaunay Triangulation (DT) each robot is able to chose the best position to be according to the position of the ball. With this feature the coach create tactical positioning for every robot in an continuous space.

In real life football, the entity most responsible for the coordination of the team is the coach. In this version of robotic soccer there is also a coach, that can be a computer outside the field making indirect decisions. For example, it is not allowed for the coach to directly order a robot, the decisions have to be taken by each player. The function of the coach is for example to change the DT tactics according to information gathered by the players.

3 The Technical Challenge

RoboCup is a competition that involves many leagues and competitions. This paper focuses on the RoboCup 2012 MSL Technical Challenge. The challenges are different every year, and this year it was composed by three active robots, one stationary goalkeeper and at least three obstacles about the same size as a robot from this league. Here is a simplified procedure for completing the challenge with respective steps:

- Robot B searches for the ball (step 1)
- Robot B passes to Robot A (step 2)
- Robot A receives the ball, dribbles for 3 meters, and passes to robot C (step 3)
- Robot C receives the ball, dribbles for 2 meters and then shoots it into the predetermined goal where the goalkeeper is standing (step 4)

For this challenge three new roles were created, Role A, Role B and Role C. Very specific instruction were given to each robot, and the ability to communicate with each other was crucial when advancing to the next step in the procedure, specially when the robots had to agree on a position to create a clear passing line avoiding the obstacles.

4 Results

By creating a challenge with three rounds and cumulative score, the RoboCup organization were rewarding consistency, for example 1 good run and 2 bad runs would only give about one third of the perfect score, to ensure teams were capable of replicating the results as many times needed, not only one lucky time.

By using shared variables though RTDB, CAMBADA robots could coordinate the passes and warn other teammates when it was their time do act. Other teams used the ball position in the field and relative distance to calculate what the next step would be, for example if the ball was is in the predefined own half of the field it probably was Robot A turn to act, and if it was on the other half of the field it would be Robot B or C turn to act. This procedure is much more risky and gives room for more errors.

CAMBADA faced serious difficulties with the vision system during the competition, as well as the other teams, because the illumination conditions were poor, and for the length of the competition serious problems affected the ball grabbing system. Nevertheless, a conservative programming allowed the team to complete the three runs with some success, finishing the challenge twice, and sending the ball out in the other after hitting an unseen obstacle. The team finished the challenge in the first place.

5 Conclusion

The ability to communicate with each other through the Real Time Database [2], the dynamic role system and the Delaunay Triangulation positioning have proven to be the best assets the CAMBADA team has to surpass its challenges. Also its distributed approach allows for a quick adaptation, either in this more familiar challenges or in other cases, like the adaptation of a CAMBADA robot to participate in the @Home competition since the RoboCup 2012.

At this point, the robots have the tools to share their knowledge with each other in an efficient way, however, using this information to make optimal decisions is the real challenge.

For future work there are two main objectives that would boost the cooperative behaviour greatly. The first would be to allow the coach to decide the best formation according to his knowledge of the world and change it during the game. In this moment there is a static formation decided at the beginning of the game. The other objective would be for the robots to coordinate passes in a more dynamic way. Currently, the destination of the pass is given by the current position of the receiver robot, but the game would be much more interesting if the robots had the ability to do through passes, for example, if two players could agree on a pass destination even before the receiver is there.

References


A platform for segmenting and characterizing polygonal networks on remotely sensed images

Nuno Benavente, José Saraiva, Lourenço Bandeira, Maura Lousada and Pedro Pina
http://planetsci-carena.weebly.com

CERENA - Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico

Abstract
This text presents an informatics platform we are developing to make a full analysis of polygonal networks, a common type of patterned ground that occurs normally on the periglacial regions of Mars and also of the Earth. These networks that, in some situations, are constituted by hundreds of thousands of polygons covering several tens of square kilometres [13, 20] can only be thoroughly analysed through automatic procedures on the remotely sensed images of very high spatial resolution. This platform is constituted by two main modules: one dedicated to the segmentation procedures, the other for extracting a set of geometric and topological features that are necessary for a full characterization of the polygonal networks.

1 Introduction
Polygonal networks are a common pattern on the surface of Mars, though they present a wide variety of types according to size and visual aspect [7, 9, 14]. The distribution of at least some of those types has been correlated with the presence of sub-superficial ice on Mars [10], and proof of that close association was established beyond any doubt in 2008 after the NASA Phoenix probe landed on the boreal plain of the planet, in a region covered with small-scale polygons, and its robotic arm found water ice a few centimetres below the surface. The similarity of the extraterrestrial features to some ice-wedge polygonal patterns on Earth is remarkable, and suggests that there is much to be learned about them by closely studying terrestrial examples. Periglacial regions on Earth where this type of patterned terrain occurs represent a good alternative for analogue field studies, namely in Canada [5, 6] and the Norwegian archipelago of Svalbard located north of the Arctic Circle [8, 24].

We have been working on the analysis of this kind of networks for a while, proposing some methodologies to make the segmentation [15] and characterization [1, 2, 23] of the polygons for automated mapping and analysis of polygonal networks on Mars [21]. We are extending this work to terrestrial examples where we can validate the mapping obtained by this method, by comparison with real data acquired on the ground [16-18, 22]. The major goal of our long-term work is advancing the knowledge about these features, namely their origin, evolution and relation with climate changes, both on Mars and on Earth [3].

The approaches we have been developing have reached a high detection rate in the segmentation procedures and have been greatly improved with fast algorithms in the characterization stage, making them suitable for interested researchers. Thus, we felt that a user-friendly software platform could be developed at the current stage of our investigation of this kind of patterned ground. This text introduces this platform and synthesizes the algorithms in which the procedures are sustained. It is organized in two main modules, one for the segmentation of the networks, the other for extracting all the features necessary for their characterization. These modules can be used sequentially or independently one from the other.

2 Segmentation module
The segmentation module is based on an approach mainly constituted by mathematical morphology operators [15]. It is organized into two main phases that can be applied to a polygonal ground independently of the type of terrain where these patterns occur.

The first phase consists of a pre-processing stage, where noise is filtered out and contours are enhanced. This goal is achieved by applying morphological connected filters, namely opening and closing by reconstruction [19].

In the second phase, the contours of the images are identified through the watershed transform [4]. This is followed by their analysis according to their relevance relatively to the minima of the adjacent basins which permits to construct an image containing information about their dynamics [12]. The adequate thresholding of this image [11] leads to obtaining the most relevant contours that correspond in these images to the cracks separating the polygons.

The interactivity of this module is helped by plotting some basic information about the image, which permits to better select and change the parameters related to the operations at each stage, for instance, the type and dimension of the filter, the type of watershed or the threshold level for the dynamics. A relatively rapid output (which is naturally dependent on the dimension of the image) is graphically produced and permits a direct verification of the adequacy of the selected parameters.

An example of the interface menu for the segmentation step is shown in Figure 1, where the input (left) and output (centre) images can be seen, as well as the image of the dynamics of the watershed crest lines (right).

3 Characterization module
The characterization module permits to obtain geometric and topological features for each individual polygon and for the whole network. The area, perimeter, axis, shape and orientation features, the number of neighbours of each individual polygon or the valence or number of edges at each vertex are properties that are computed. Moreover, the existence of a correlation between geometry and topology on polygonal networks of widely diverse nature has been verified and expressed in a number of laws generally known by the names of their proponents, such as Lewis, Desch and Aboav-Weaire [23]; the necessary parameters to verify the correlations between geometrical (area or perimeter) and topological features (number of neighbours of a polygon or number of neighbours of the neighbours of a given polygon) and the correlations themselves are also computed.

Anyhow, the classic approaches are not able to deal with large amount of polygons, so a novel approach to extract topological features of extensive polygonal networks constituted by several thousands of polygons, based on a multi-layer strategy inspired by the four-colour problem, was obtained [2]. The major improvement of the algorithm consists of the distribution of polygons by layers in such a way that adjacent polygons cannot coexist on any given layer; this is followed by a global analysis of each layer to extract topological features. This novel approach can be indistinctively applied to any kind of tri and tetravalent network (presenting respectively three and four polygons at each vertex); its computational performance is extremely favourable when compared with previous approaches to this problem.

An example of the interface menu of the platform concerning the characterization phase is presented in Figure 2, where some features of the detected network are presented.
4 Conclusion

This platform permits an easy and fast analysis of relatively large polygonal networks of planetary periglacial regions observed on remotely sensed images. Its conceptual basis allows an interactive search by the user for the best parameterization of the different procedures of the algorithm.

The modularity of the platform also permits the segmentation and characterization phases to be used sequentially or independently one from the other.

Diverse types of output formats (image or text) are also available.

Acknowledgements

This work is being developed in the frame of the projects ANAPOLIS (PTDC/CTE-SPA/099041/2008) and ANIMAR (PTDC/CTE-SPA/110909/2009) funded by Fundação para a Ciência e a Tecnologia.

References


Unravelling the molecular determinants behind the tumour spectrum associated to TSG: CDH1 and HDGC as model

Patricia Carneiro
Patricia Oliveira
Joana Figueiredo
Carla Oliveira
Raquel Seruca, PI

IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto; Medical Faculty of the University of Porto, Portugal
http://www.ipatimup.pt/Site/Home.aspx

Abstract

Tumour suppressor genes (TSGs) code for proteins for which complete loss of function increases the risk of a normal cell to be transformed into a neoplastic one. Germline mutations in TSGs predispose to hereditary cancer in specific tissues, despite that these mutations are ubiquitously expressed. This is the case for CDH1 germline mutation in hereditary diffuse gastric cancer syndrome (HDGC), with thirty to forty percent of HDGC suspected families harbouring germline CDH1 mutations (HDGC; OMIM137215). Carriage of the abnormal CDH1 gene confers more than an 80% lifetime risk of developing gastric cancer, upon somatic inactivation of the remaining CDH1 wild type allele. However, the tissue specific molecular determinants behind the predominant increased risk for stomach cancer in CDH1 germline mutation carriers remain to be uncovered. In order to unravel this biological enigma we propose a multidisciplinary strategy using bioinformatic approaches as well candidate gene expression and functional analysis in diverse tissue types, namely gastric cancer.

Introduction

Germline mutations in TSGs predispose to hereditary cancer associated to a specific tumour type. In fact, loss of function of specific TSG cause a different cancer associated syndrome (affecting a different number and type of tissues) despite their ubiquitous expression. BRCA1 and BRCA2 are two good examples of TSGs that develop a tissue-restricted panel of tumours: BRCA1 and BRCA2 loss of function lead to hereditary breast and breast-ovarian cancer, respectively. To address this challenging issue of tumour spectrum associated with germline mutations of specific TSGs, we took advantage of the extensive and solid experience of the group in Hereditary Diffuse Gastric Cancer (HDGC) research and CDH1 (E-cadherin coding gene) germline mutations as model. CDH1 germline mutations heterozygous carriers are at risk of developing, specifically, diffuse carcinomas in stomach, despite the ubiquitous expression and function of E-cadherin in all epithelial tissues. Lobular breast cancer has been described as a rare entity in HDGC. No other types of tumours in other organs have been clearly associated to HDGC. Therefore, the clinical criteria to suspect of HDGC defined by the IGCLC were: (1) two or more documented cases of diffuse gastric cancer in first/second degree relatives, with at least one diagnosed before the age of 50; or (2) three or more cases of documented diffuse gastric cancer in first/second degree relatives, independently of age.

However, the reasons why CDH1 germline mutation carriers are at risk of developing this particular tumour type in this specific organ (diffuse stomach carcinoma) remain to be answered. One hypothesis is the putative implications of gastric tissue-specificity patterns of expression associated to E-cadherin and their contribution to the development of HDGC. This is a challenging enigma. In fact, nothing is known about the molecular determinants leading to the specific increased risk of stomach cancer and not other epithelial tumours in E-cadherin mediated–HDGC in particular, or diverse tumour spectrum associated to other to TSGs, in general.

Our working hypothesis is that CDH1 biallelic inactivation is only tolerated in gastric epithelium and not in other epithelia due to a favourable gastric-specific expression program that allows cell-autonomous survival as a first step towards tumourigenesis. Such cells will then become more prone to acquire de novo alterations, as a second step, leading to cell transformation. Understanding the pattern of expression associated to this tissue-specific tolerance of CDH1 inactivation in stomach will be of use for the design of innovative therapeutics for carcinomas with E-cadherin deregulation (hereditary and sporadic).

Problem Formulation

As long term goal, we aim at identifying the patterns of genetic expression associated to stomach-specific survival genes upon E-cadherin loss plus those genes associated to increased transformation. At this stage of research, we concentrated on the identification of the gastric-specific molecular program underlying the ability of gastric cells to overcome E-cadherin loss of function in stomach in contrast to other epithelia. In fact, we aim to identify the gastric-specific molecular program underlying the ability of gastric cells to overcome E-cadherin mediated apoptosis (specific cell death program).

Accordingly, we demonstrated that gastric, breast and colon cancer derived cells react differentially to E-cadherin depletion, with gastric derived cancer cells displaying higher resistance to pro-apoptotic stimulus when compared to breast and colon cancer derived cell lines, which died massively by apoptosis. Further, we showed that GC cells silenced for E-cadherin are more resistant to pro-apoptotic stimulus than control cells, in a Notch signalling-dependent manner.

Strategy

We propose to identify the gastric-specific molecular program underlying the ability of gastric cells to overcome E-cadherin mediated apoptosis, by addressing the following workplan:

• Uncover human gastric-specific genes using a bioinformatic/data-mining approach of public available data on microarray experiments;
• Analyze the expression profile of gastric-specific genes involved in survival/apoptosis in RNA derived from normal stomach, colon, breast and other epithelial tissues in order to determine the genetic make-up of each tissue type;
• Determine the pattern of expression of several candidate stomach-specific genes in a series of cell lines and primary tumours from diverse epithelial origin (gastric, breast and colorectal) to determine putative gastric cancer related genes;
• Analyze the response of normal and cancer epithelial cells from stomach to E-cadherin loss in terms of cell survival/apoptosis;
• The functional relevance of candidate gastric E-cadherin-related survival keepers in stomach will be evaluated using RNAi; to . Colon and breast cancer cells will be used for comparison.

Experimental Results

Public databases such as GEO and ArrayExpress encompass a large and freely available source of information on gene expression pattern in different tissues. We have extracted data from these databases concerning 3 published microarray experiments performed using up to 35 distinct normal tissue types in order to uncover a list of stomach-specific genes. For a given gene to be classified as stomach-specific, its normalized expression needed to be: (1) higher than the average across all tissues plus 3 times the standard deviation value; or (2) higher than 2 times the expression obtained in the second tissue with highest expression. Each dataset was analyzed separately given that different gene expression platforms and methodologies were used. We have also produced a fourth dataset of stomach-specific genes using the web resource TiGER (tissue-specific gene expression and regulation), which stores the results derived from a massive computational analysis, summarizing several large scale data sets, searching for, among other features, tissue-specific gene expression in a variety of human tissues.

For each of the thus-obtained four datasets of stomach-specific genes, we performed functional annotation clustering using DAVID (The Database for Annotation, Visualization and Integrated Discovery), which performs
a functional annotation enrichment analysis of a given set of genes. In this we have validated the bioinformatic analysis by RT-PCR and case, stomach specific genes, by accessing a variety of public confirmed expression of genes to be confined to stomach tissue RNA bioinformatic databases. We ascertained in which significant gene (Figure 1B). A set of genes displayed retained expression in gastric ontology or functional categories are these genes involved. This analysis cancer cell lines (diffuse type), namely MUC1, S100P and CTSE was carried out for each of the 4 datasets of stomach-specific genes. (Figure 1C). Overall, we observed that stomach-specific genes showed a significant involvement with development, cell proliferation, growth and apoptosis.

in 3 of the datasets up to 27% of stomach-specific genes are related to cell proliferation, growth and apoptosis while in 2 of the datasets up to 29% of stomach-specific genes are significantly related to development. A final and more restricted stomach-specific gene list was compiled based on the 4 datasets by assessing which genes were present in at least two of those datasets. We obtained 25 common stomach-specific genes, which were then matched to state of the art literature aiming at identifying which of them (from normal tissue) may be relevant to cell survival/apoptosis and gastric cancer. This revealed that 15 out of these 25 stomach-specific genes have been reported as displaying a cell survival/apoptosis relationship with gastric cancer (Figure 1A).

Conclusions

Our preliminary data indicates that cumulative activation of gastric-specific genes may be underlying the ability of gastric cells to overcome E-cadherin mediated apoptosis in contrast to other tissues. However, we believe that state of the art data-mining approaches should enable us to efficiently identify the molecular determinants of E-cadherin mediated HDGC tumour spectrum. For this purpose, we will next analyse the stomach-specific genes identified bioinformatically aiming at uncovering both structural and/or functional patterns, namely the existence of common domains. The identification of genomic/proteomic patterns could allow the identification of other stomach-specific genes and, more importantly, provide hints regarding the relevance of such genes in a stomach context and their putative link to cell survival/apoptosis. Analysis of the expression profile of gastric cancer samples (already available in public databases) will further enable us to identify which genes must be altered for tumourigenesis to occur. Integration of these approaches along with the functional pattern recognition previously proposed will allow us to dissect the tumour spectrum underlying gastric tumourigenesis.

References


Acknowledgments

We thank the Portuguese Foundation for Science and Technology (FCT) for funding through the Project PTDC/SAU-ONC/110294/2009 (PI: R Seruca).
Characterization of the intronic portion of cadherin superfamily members

Patrícia Oliveira
Remo Sanges
David Huntsman
Elia Stupka
Raquel Seruca
Carla Oliveira, PI

IPATMUP - Institute of Molecular Pathology and Immunology of the University of Porto; Stazione Zoologica Anton Dohrn, Napoli, Italy; British Columbia Cancer Agency, Vancouver, Canada; Center for Translational Genomics and Bioinformatics, San Rafaele Scientific Institute, Milan, Italy; Medical Faculty of the University of Porto, Portugal

http://www.ipatmup.pt/Site/Home.aspx

Abstract
Cadherins are cell-cell adhesion proteins essential for maintenance of tissue architecture and integrity and their impairment is often associated with human cancer. Knowledge regarding regulatory mechanisms associated with cadherin misexpression in cancer is scarce. Specific features of the intronic-structure and intronic-based regulatory mechanisms in the cadherin superfamily are unidentiﬁed. This study aims at systematically characterizing the intronic portion of cadherin superfamily members and the identiﬁcation of intronic regions constituting putative targets/trigger of regulation, using a bioinformatic approach and biological data mining. Our study demonstrates that the cadherin superfamily of genes harbours speciﬁc characteristics in comparison to all non-cadherin genes, both from the genomic and transcriptional standpoints. Cadherin superfamily genes display higher average total intron number and signiﬁcantly longer introns than other genes and across the entire vertebrate lineage. Moreover, in the human genome, we observed an uncommon high frequency of MIR and MaLR regulatory-associated repetitive elements at 5′-located introns, concomitantly with increased de novo intronic transcription. Using this approach, we identiﬁed cadherin intronic-speciﬁc sites that may constitute novel targets/trigger of cadherin superfamily expression regulation.

These ﬁndings pinpoint the need to identify mechanisms affecting particularly MIR and MaLR elements located in introns 2 and 3 of human cadherin genes, possibly important in the expression modulation of this superfamily in homeostasis and cancer.

Introduction
Cadherins are transmembrane glycoproteins involved in biological functions from tissue morphogenesis to cancer. All cadherin proteins share one or more copies of a 110 residue extracellular peptide (cadherin repeat-EC), responsible for mediating calcium dependent homophilic/heterophilic cell-to-cell adhesion. Many cadherin superfamily members have been proved or suggested to work as tumour suppressor genes and oncogenes in different cancer contexts. Changes affecting cadherin expression are of particular relevance in epithelial cancers, which constitute approximately 80-90% of all human cancers. In this context, the non-homeostatic loss of cellular adhesion is frequently the master trigger for invasion and metastatization. E-cadherin (OMIM*192090), a classical tumour suppressor gene, possesses anti-invasive and anti-metastatic properties and the clinical turning point in carcinoma progression and metastasis is mediated by its disruption in 90% of all epithelial cancers. Although classical gene inactivation (mutation, gene loss and promoter hypermethylation) and transcriptional and post-transcriptional mechanisms (transcriptional repression, RNA and protein quality control) hamper normal E-cadherin expression and function, these phenomena are insufﬁcient to explain E-cadherin impairment both in development and overall tumour progression. Regulation by non-coding RNAs (microRNAs), alternative transcripts (antisense transcripts) and alternative translated isoforms (antagonistic isoforms) have recently emerged as a new layer to explain cadherin impairment both in development and overall tumour progression. Regulation by non-coding RNAs (microRNAs), alternative transcripts (antisense transcripts) and alternative translated isoforms (antagonistic isoforms) have recently emerged as a new layer to explain cadherin impairment both in development and overall tumour progression.

Regardless of the numerous observations in cadherin superfamily members’ mis-expression in cancer and other diseases, the mechanisms that control these effects are vastly unknown and increasing evidence attribute protein functional impairment to regulation by or at non-protein coding intronic and intergenic sequences.

Aim
The aim of this study was to systematically characterize the intronic portion of cadherin superfamily members, in order to identify regions constituting putative targets of regulation. Using a bioinformatic approach, mining current gene annotation data, as well as biological data from the ENCODE project, we investigated genomic and transcriptional intronic-related features of this gene superfamily.

Experimental Results
1. Selection of cadherin superfamily members
To characterize the genetic architecture of as many members as possible belonging to the cadherin superfamily, we generated criteria that could successfully gather all genes within this superfamily, using the Ensembl database. We combined literature data and protein domain databases (InterPro) to collect genes that coded for protein sequences currently classiﬁed as cadherins, resulting in seven distinct cadherin-related InterPro domains. Selected genes were further curated to assure that only protein coding genes were selected and included. All protein-coding genes in each genome that were not selected for the cadherins’ dataset were grouped in a control dataset named non-cadherins (for example, human cadherins = 104, mouse cadherins = 91 genes). All data was extracted from the Ensembl database using private PERL scripts and data was stored in private MySQL databases. Enquiries to the databases was done using private PERL scripts and statistic analysis was done using the program R.

2. Despite no sequence conservation, cadherins display a higher and wider total intron number
We started by analysing the conservation of the cadherin intronic sequences (Genomic Evolutionary Rate Proﬁling Constrained elements) across 12 mammalian species in comparison to non-cadherins. Novel datasets of randomly selected introns from within non-cadherin genes and distribution of cadherin/non-cadherin introns into speciﬁc length classes (each length class possessed the same total nucleotide number, corresponding to 1/3 of the sum of the length of all introns extracted from Ensembl for each species) were some of the approaches undertook to prevent any unforeseen bias in the analysis. No signiﬁcant differences in overall intronic sequence conservation were observed when comparing cadherin with non-cadherin introns. Next, and using the total intron number as a measure of gene architecture, we have observed that overall, the cadherin superfamily of genes presented a higher average intron number than non-cadherin genes in all species analysed (>11 introns for cadherins and <9 for non-cadherins, p<0.05, Figure 1). The analysis of the distribution of total intron number revealed that while non-cadherins displayed a single and highly enriched peak corresponding to genes with less than 10 introns, the intron number distribution of cadherin genes resulted in four distinct peaks (human and mouse genomes): peak 1 corresponded to <10 introns; peak 2 corresponded to 12-20 introns; peak 3 corresponded to 30-40 introns, and; peak 4 corresponded to 50-80 introns (Figure 1). To understand whether the wider range of total intron number resulted mainly from the comparison of a single family of genes against all genes in the genome, we compared the total intron number in cadherins to that of other human gene families/sets both related or not with cadherins’ top biological function, adhesion. By comparing intron number distribution with the Inter-Quartile Ranges-IQRs (which...
measures data dispersion) from cadherin genes with those of the remaining families of genes, we observed that cadherin genes displayed a wider total intron number distribution in comparison to all other families/datasets, supporting our previously observation for comparison with non-cadherin genes, and proved that cadherins are in fact a particular family of genes in terms of total intron number distribution.

3. Cadherin genes have significantly longer introns than other genes in the vertebrate lineage

The assessment of cadherin genetic structure revealed that cadherin family members such as CDH1, CDH2 and CDH3, presented unusually large introns. Given the growing evidence of functional regulatory features located within introns\(^2\), we analysed whether the cadherin superfamily of genes displayed significantly longer introns than other genes in the genomes of Homo sapiens, Mus musculus and Danio rerio. Cadherin and non-cadherin introns were separated according to their length (three length classes) and their status as the longest intron from each gene or as non-longest.

This comparison showed a significant enrichment of all cadherin introns (both longest and non-longest) in length class 3 for all three genomes (\(p<0.05\), Figure 2 for Homo sapiens). All non-cadherin introns analysed in turn, were significantly more present in length class 1 (\(p<0.05\)). These data indicate that cadherin introns are significantly longer than those of non-cadherin genes (regardless of the longest/non-longest status) and that this feature is conserved, being therefore potentially relevant to the entire vertebrate lineage.

4. Human cadherin introns exhibit increased MIR and MaLR frequency and simultaneous increased intronic transcription

Prior reported studies indicated that in eukaryotes, first introns (positioned closer to the 5'-end of genes) tend to be longer, and often harbour relevant regulatory elements essential for gene expression control. Our analysis observed that both cadherin and non-cadherin introns follow such trend. We then tested whether 5'-located cadherin introns would accommodate important regulatory-associated annotated features, and determined the frequency of such features within cadherin introns. We analysed the presence of several elements, however, only for the specific DNA repetitive sequences from the families of Short Interspersed Nuclear Elements (SINE) and Long Terminal Repeats (LTR), as differences observed between cadherin and non-cadherin introns. SINE elements, in particular Alu and MIR (Mammalian-wide Interspersed Repeats) and LTR elements, in particular MaLR (Mammalian-Wide Interspersed Repeats), were found to be present in a significantly distinct manner in cadherin introns: 1) Alu elements were significantly less frequent in cadherin introns than in non-cadherin introns (\(p<0.05\)); while MIR and MaLR elements were significantly more frequent in cadherin introns (\(p<0.05\)) in comparison to non-cadherin introns.

MIR and MaLR repeat elements are involved in genome novelty by alternative regulation phenomena as well as by promoting exonization. Therefore, we next analysed whether the unusual frequency of Alu, MIR and MaLR elements within cadherin introns was correlated with differential intronic transcription arising from cadherin introns. To assess cadherin intronic transcription, we mined the data from the pilot stage of the ENCODE project, in particular data obtained by 5' cap analysis gene expression performed by Carninci et al. This technique allowed for the detection of new TSS and we focused on the data obtained using RNA extracted from cytosolic and nucleic fractions of the lymphoblastic human normal cell line GM12878 to assess transcription\(^4\).

Moreover, given the prior observations on cadherin introns’ increased length and differential repeat frequency we realanalysed these features in concomitance with the intronic transcription data, in terms of intron positioning. We verified that cadherin introns at positions 2 and 3 displayed a significant TSS enrichment in the cytosolic fraction of the cell line GM12878, arising from both RNA strands (\(p\) ranging from 1.51E-04 to 4.01E-02, Figure 3). Taken together, a correlation was observed for cadherin introns found at positions 2 and 3 for which cadherin introns were significantly longer, carried increased frequency of MIR and MaLR elements and increased levels of novel transcription initiation than non-cadherin introns (Figure 3).

Conclusions

Our study demonstrates that the cadherin superfamily of genes harbours highly specific characteristics from the genomic and transcriptional standpoints, namely high frequency of specific repetitive elements within cadherin 5'-located long introns combined with an unusual frequency of novel transcription initiation. These findings lay the ground for discovering novel areas important in fine-tuning the expression of this gene family as well as intronic-based regulatory mechanisms, particularly in introns 2 and 3, important for expression of the cadherin superfamily of genes in biological events such as cadherin switching or cadherin gene loss/functional impairment in homeostasis and disease.

References


Acknowledgments

We thank the Portuguese Foundation for Science and Technology (FCT) for funding through the Projects PTDC/SAU-GMG/72168/2006, PTDC/SAU-GMG/110785/2009.
Abstract

This paper describes a new methodology to improve band detection in Thin-Layer Chromatography (TLC) images. The storage of the biological samples to be analysed by TLC is usually done via plastic containers. Filter paper is an alternative that allows reduced costs and higher portability, at the cost of consequences in the image analysis stage due to a lane background alteration. In order to overcome this problem, a negative control lane is generated in every chromatographic plate. After the pre-processing and lane detection stages, a one-dimensional intensity profile is used for integrating lane information and the background influence is removed with the help of the Continuous Wavelet Transform (CWT) decomposition. The proposed method was tested in the classification of 312 bands (175 true bands) and presented better results than the ones obtained using the original lane profile.

1 Introduction

The objective of this work is the development of a methodology to improve the automatic analysis of digital images of Thin-Layer Chromatography (TLC) [1] plates. The biological samples to be analysed by TLC are usually stored and transported in plastic containers. Instead of this conventional sample storage, filter paper can be used since it presents the main advantages of reduced costs and higher portability. However, the use of filter paper has consequences in the image analysis stage due to the diminished concentration of the compounds, as well as the inclusion of additional substances in the sample, which causes a lane background alteration when compared to what is expected without the use of filter paper. In order to use image analysis techniques for automating the detection of the main compounds that are present in a sample, it is important to minimize the background influence. For this purpose, the material resulting from the application of the extraction process to plain filter paper (without the addition of any kind of biological material) will be included in every TLC plate for generating a negative control lane. As this negative control lane will only contain the result of chromatographic separation of the filter paper components, the associated lane pattern can be used for removing the background of all the other sample lanes.

Fig. 1 presents a digital image of a TLC plate containing several vertical lanes each one corresponding to the development of a sample. The image pattern of each lane is formed by horizontal bands associated with the specific sample compounds. The image clearly illustrates the influence of filter paper usage. The leftmost lane (limited by the blue vertical lines) is the negative control lane; the samples in following 10 lanes were extracted from filter paper, while all the other lanes on the right resulted from the chromatographic development of samples stored using the conventional containers. The rightmost lane, delimited by the black lines in the figure, contains a standard combination of known compounds (standards lane). The two sample lanes delimited by the red lines were obtained using exactly the same quantity of a specific compound (Gb3). As can be observed, besides the distinctive background, the bands corresponding to Gb3 also have a completely different contrast.

2 Methodology

Several image and signal processing operations are applied to a raw TLC image to extract a set of lanes. A short description of these techniques can be found in subsection 2.1.

In order to accomplish a satisfactory attenuation of the background present in the lanes, various methodologies were proposed and implemented by the authors. The one herein presented is based on the continuous wavelet transform decomposition of the lane profile.

2.1 Pre-processing and lane detection

The processing flow applied to the original image of a TLC plate is illustrated in the block diagram of Fig. 2. Two distinct regions can usually be found in a typical image: an internal region of interest (ROI) and an external border that has no relevant information for the image analysis process. The ROI is detected using a clustering based algorithm described in [2] and afterwards converted to grey scale. The ROI background is estimated and removed using a closing morphological operator with a large square structuring element. The image is then projected on the direction perpendicular to lane development (vertical projection) to integrate all the image data into an intensity profile. As during background removal the image is inverted, the intensity profile presents maxima regions where the original image is darker (lane zones) and minima regions where this image is lighter (empty zones). The intensity profile is then smoothed using a Savitzky-Golay filter, and lane detection proceeds as described in [3].

Figure 2: Main steps of the methodology used for obtaining a set of lanes ready for classification.

2.2 Lane background removal

The information of each image lane is integrated into an intensity profile by projection along the sample development direction. The profiles are afterwards smoothed using a procedure based on the Continuous Wavelet Transform (CWT) decomposition. The CWT performs signal decomposition through a set of orthonormal functions generated by the scaling and translation of two functions: the mother wavelet and the daughter wavelet. As a result of the wavelet transform, a set of coefficients hierarchically organized is produced, each one associated with a spatial position and a frequency component (scale) of the original profile [4].

The scales in the CWT are not constrained and the wavelets are non-orthogonal. This property, while making CWT redundant, provides a finely detailed description of a signal in terms of both time and frequency [5]. These characteristics allow an accurate selection of scales that will be important in the smoothing of the intensity profiles, since we pretend to adapt the level of smoothing to the type of lane (sample or
negative control). The smoothing level can be controlled by selecting a cutoff in the scale domain and by forcing all the coefficients at lower scales to zero. A cutoff value at a high scale (low frequency) will ensure greater smoothing than a cutoff at a low value (high frequency).

Fig. 3 depicts the images of two lanes (top) and their wavelet decomposition (bottom), with a negative control lane shown on the left and a sample lane on the right. This representation helps to perceive which range of scales contain the most important coefficients for each type of lane. From the analysis of several examples, it was possible to conclude that the majority of the high amplitude coefficients in the negative control lane (Fig. 3a) are located at high scales (low frequencies). In this case, it is advantageous to use a high cutoff value to remove most of the noise and retain just the low frequency components which are associated with the lane background that is expected to occur also in all the other lanes. In what concerns the sample lanes (Fig. 3b), there are relevant low scale (high frequency) coefficients related with the presence of bands. For this type of lanes, the cutoff value should be lower than the one used for the negative control lanes. The cutoff values were estimated based on a randomly selected set of negative control and sample lanes.

Figure 3: a. Negative control lane b. Sample lane.

For background removal, the smoothed profile of the image negative control lane is subtracted to the smoothed sample lane profile. Fig. 4 shows the process of background removal for the lane on the right of Fig. 3. Fig. 4a and Fig. 4b present the original (top) and smoothed (bottom) profiles for the negative control lane and the sample lane, respectively. The profile resulting from background removal is presented in Fig. 4c, while Fig. 4d shows the reconstructed lane image obtained by profile replication.

Figure 4: a Negative control lane original (top) and smoothed (bottom) b. Sample lane original (top) and smoothed (bottom) profiles. c. Result of background removal. d. Reconstructed lane image.

3 Results

The proposed approach was tested in a total of 78 lanes, each one with 4 regions where a band could exist (312 potential bands). The lanes were previously examined by a specialist for marking the true bands (175).

For band detection, the lane profile in each of the 4 potential band positions is analyzed for calculating band densities. The two local profile minima that define the band width are found, and the area limited by the profile and the line connecting these two minima is used for band area calculation. Band density is obtained as the ratio between band area and the line connecting these two minima is used for band profile minima that define the band width are found, and the area limited by the profile and the line connecting these two minima is used for band area calculation. Band density is obtained as the ratio between band area and the line connecting these two minima is used for band profile minima that define the band width are found, and the area limited by the profile and the line connecting these two minima is used for band area calculation. Band density is obtained as the ratio between band area and the line connecting these two minima is used for band profile minima that define the band width are found, and the area limited by the profile and the line connecting these two minima is used for band area calculation. Band density is obtained as the ratio between band area and the line connecting these two minima is used for band area calculation.

In order to evaluate the lane background removal process, the results of band detection obtained with and without background were used to obtain the two ROC curves shown in Fig. 6, which were obtained by a variation of the threshold value used for band detection.

![ROC curves for band detection obtained with and without lane background removal.](image)

4 Conclusions

The additional problems found in the analysis of chromatography images when filter paper is used to store and transport the biological samples were addressed by including a negative control lane in every TLC plate. The integration of lane information into a one dimensional profile along with an efficient smoothing proved to be effective for the estimation and removal of the lane background influence.

As future work we intend to improve the methodology for obtaining the intensity profile by a previous compensation of some distortions that occurs in the bands, such as bending and smiling effects. We also intend to measure other features besides band density to improve the robustness of the band detection process.

Acknowledgements

The authors would like to thank Prof. Maria Clara Sá-Miranda from IBMC for providing the datasets used in this work. 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 PTDC/SALBEB/100875/2008.

References


Abstract
This paper has a two-fold intention: i) to describe our global research proposal; ii) to present our on-going work towards the understanding of semantic human behavior in a retail complex scenario, which ultimate goal is to build a behavioral semantic model that analyzes shopper’s non-verbal features individually and in group, their relations with the scene, and their needs for assistance support.

1 Introduction
Human activity understanding is an important and growing area of computer vision research. It is an area that explores automatic reasoning mechanisms to express human activities by high-level semantics, which can be composed for multiple atomic actions, entities relationships, and context information. However, until the recent years, this area was predominant influenced by recognition methodologies that just analyze human body taxonomies. Thanks to psychological, sociological and cognitive studies, increasing of computational hardware power, and to the advance of parallel areas in computer vision such as object recognition, scene layout recovery, and scene interpretation, most recent works are starting to merge mutual context with the aim to improve human activity understanding.

At the level of human-human interactions, Jin et al. [7] used context-free grammar to recognize interactions, based on production rules governed by semantic spatial relationships between individuals obtained from multi-tracking algorithms. Choi et al. [2] recognized collective human activities by modeling common and repetitive configuration behaviors of individuals, and their relationships distributions over time and space.

Considering relations between human-space, long-term observations of moving objects in the scene allow to build semantic scene models from the spatial distribution of trajectories. Pusiol et al. [10] designed an intermediate layer composed of Primitive Events descriptors, to cluster motion trajectories by individual segments of slow points in trajectories and extract meaningful transition between topological slow regions.

The work of Gupta et al. [4] explored relations between scene and objects within it, by building a physical representation that models objects as volumetric entities, and their relationships to describe the 3D structure and mechanical layout of the scene.

This section highlights an important change of research paradigm that states that the understanding of 3D activity should be treated as one complex problem since the beginning, rather than divided by a number of independent detached problems. Next, we present an overview of our research proposal in section 2, followed by a preliminary work on a real application scenario, section 3. Finally, we formulate our conclusions and directions for future work in section 4.

2 Global research proposal
Our proposal refers to an unsupervised perceptual framework that can help us to understand how to identify and classify high-level human activity in indoor environments at multiple set of temporal scales under an specific application context. We are interested on integrating microscale representation with context modeling to infer characterization of human activity.

Our statement considers: i) Active-Learning of Concept-Characterizations: [5] proved that activity classes could be modeled as combinations of their local sequential features. Our aim is to add perceptual information about the elements of activity dynamics at any stage to improve inference of concepts and categories; ii) Object and Movement Space Mapping: action primitives [8] are used for both recognition and synthesis. We believe that their associated grammatical description could be integrated as activity descriptors; iii) Discriminative Context Modeling: [11] showed that incorporation of shape, appearance and context is an efficient method for image categorization and segmentation. We believe that this approach could benefit from semantic and spatial relations; iv) 3D Spatial Structure: [12] proved that spatial structure reflects the functionality of the location and help to suggest scene category. We believe in a holistic approach for recognize functional scene layout.

Our novelty resides in an unified approach that deals with the interpretation of objects, scene, actions, and their mutual contextual constraints to improve action classification, scene context categorization, and semantic inferring. We believe that application goals will largely benefit from this perceptual framework, and as an example, we present next, in section 3, a brief description of our first approach, based on motion characteristics, to extract human behaviors on shopping scenario.

3 Human behavior in retail scenario
Understanding customers behaviors and their purchase decision processes in an automatic way is an inestimable commercial advantage for the retail market. These kind of applications have been facing a growing demand; However, there is not any fully automated system for customer behavior analysis commercially available.

Our major contribution is in the comparative performance of two systems to compute global motion field for path learning in a very challenging scenario: a video with small resolution, low frame rate (1 fps), and uncontrolled camera deployment process of a real cluttering shopping store with heavy crowds, occluded bodies, and reflective surfaces, where customers move, individually and in group, randomly in various directions, and whose appearance models changes abruptly from frame to frame.

The traditional approach for motion analysis consists of detecting objects, tracking them, and analyze their tracks for event/activity detection. This standard processing does not work well on high density scenes with cluttered environment since reliable trajectories of objects can not be obtained. To solve these cases, global motion flow field is used to learn typical motion patterns. Under these constraints and considering motion field acquired from optical flow algorithms, we implemented and evaluated two approaches: grid-based global dominant motion flow method, and kernel-based sink-seeking method. These methods belong to the last processing step of the baseline framework composed by the following steps: sampling, motion flow fields, matching, and motion flow tracking.

The sampling process is responsible to extract a set of relevant points for frame matching, and can be subdivided into two types: i) dense, ii) sparse. Normally, for image classification and action recognition dense sampling performs better than sparse sampling, since it enables local reasoning from motion similarities, and introduces spatial regularity constraints in the clustering method. However, we verified empirically that for our dataset sparse sampling is preferred, not only for computation effort when generating the global motion trajectories, but also because dense sampling introduces noisy points from the cluttered background, and does not increase discriminative value for trajectories.

The motion flow fields are obtained from existing optical flow algorithms that consider frame-to-frame analysis or larger spatio-temporal displacements. This step is computed independently, and the resulting motion flow map is used in the matching step process, in conjunction with the sampling points. Since our dataset has very hard demanding conditions, we considered two types of optical flow algorithms for motion estimation: i) short-classical, ii) large displacement. We verified that this step
largely influences the final output. From several algorithms tested, we highlight two of them: the short-classical Farnebäck [3], and the descriptor matching in variational model (LDOF) [1]. Both of them present good results, however the latter is an offline method that permits the extraction of smoother and longer trajectories, and the former is computed on real time but introduces noise on areas with large appearance variations.

The matching step consists in the tracking of each sampling point \( P_i = (x_i, y_i) \) from frame \( f \) to next frame \( f + 1 \), to obtain point \( P_{i+1} = (x_{i+1}, y_{i+1}) \). These two points form a flow vector at a specific pixel location. The adopted approximation is a median filtering kernel, that performs better than the bilinear interpolation and removes impulse noise, preserves edges, and smooths points in dense optical flow fields.

The motion flow tracking step has a common pipeline to extract meaningful flow vectors. The image region is divided by an equally spaced grid, and each cell contains the flow vectors that lay inside it. On each cell is applied a two-step hierarchical clustering approach with a two-fold purpose: to reduce the number of flow vectors, still maintaining the geometric structure of the flow field, and to obtain the local dominant motion flows. The first clustering step considers a full-orientation histogram with 8 bins to express the orientation groups. The groups with weight above the histogram’s mean are taken into account for next clustering step. The second step implies a spatial clustering on each orientation group. At the end, there are several clusters for each orientation group that are ordered in a descendent-weighted way, considering the number of flow vectors that belong to them. These groups represent the local dominant flows.

3.1 Results

The two implemented algorithms of the final processing step are: i) **Grid-based Global Motion Flow**: computes the global dominant motion flows derived from local dominant motion flows in neighborhood, and is based on Ozturk et al. [9] approach. Each cell has a depth-step for looking for neighbors, which are defined as the regions that are in the direction of the current local flow. The scanning process considers a first iteration, where the current dominant flow just search for neighbors that belong to the same orientation group, and in case of no returning the subsequent iteration consider the closest flow in the valid neighborhood to connect with it, which are the adjacent orientation groups. If no neighbor flow is obtained in the valid depth-step, any orientation is considered to keep continuity and permit abrupt flow orientation changes; ii) **Kernel-based Sink-Seeking**: follows the work of Hu et al. [6], and considers a sink-seeking process, which is a sliding-window-based technique that uses a continuous kernel-based estimator to obtain global motion paths. It incorporates neighborhood flow motion information to obtain the representative states of the global motion path. The relation between next and previous states is linear, and the next location depends on position and flow motion of neighboring points. The neighborhood is defined by the flow vectors that lay inside the kernel window, and whose angle difference against the current angle sink state is below a certain threshold. The linking between consecutive sink-seeking states gives the global motion flow. Both algorithms extract meaningful motion trajectories indicating the global motion flows in cluttered regions, however, they perform differently. A global comparison: the kernel-based method produces smoother but shorter trajectories than the grid-based, probably since the appearance changes a lot from frame to frame, the kernel window does not accept flow vectors with large opposite directions, so the sink-seeking process maintains a coherent flow, and since it is a sliding window technique, its overlap permits to build smoother trajectories; the kernel-based method is much slower, since the sink-seeking process is repeated for each flow vector; the kernel-based method produces much more trajectories than grid-based, so it requires a sparse sampling and just one local dominant flow vector by cell to run on real time; the grid-based method is more sensitive to noise and does not keep an smooth flow.

Individually each algorithm has a combinatorial number of factors that affects the final performance. The grid-based method’s results are affected by: i) the pair-wise metric used to choose the next local dominant flow from the returned neighborhood’s local flows, which can be just spatial, or a weighted additive distance; ii) the neighborhood, which has a depth-step factor that affects the search radius among cells, and the similarity metric between current and next orientation group, which can be equal, similar, or other, provoking the continuity and/or discontinuity of flow. In turns, the kernel-based method presents other factors: i) the bandwidth size, which largely affects the sink-seeking process and has three important steps that influence results: size initialization, size restart on each vector flow, and size updating; ii) the acceptance angle, which is a threshold to filter the flow vectors that participate on the kernel window, the greater more probabilities to have opposite vector flows to push current trajectory to other directions.

Ground truth data was obtained from manual tracking annotation, for each person two reference points were considered: head and center of mass. A parallel spectral clustering, with K-means on its final stage, was applied on distance matrices of trajectories from ground truth data, and from result data of both algorithms. After obtaining the clusters, we adopt a partition-and-group framework to extract the most representative trajectories. These results are presented on figure 1.

4 Conclusions

In this paper we briefly describe our research proposal justified for its relevance on scientific community, and explain its purpose and foundations. A real case scenario for understanding shopping behavior is summarized and its qualitative results are discussed. This reflects a work-in-progress, whose next step will be the validation of exact metrics to evaluate the performance on computing trajectories based on motion flow, a trajectory encoding scheme to estimate space layout and semantic local regions, and implement a tracking social model to infer human-human relationships. The intention is to close the loop and use mutual context to integrate human, objects, and space relationships together.

References

E-cadherin spatial characterization with radial distribution profile for mutation detection.

Sofia Esménio¹
sof.esm@gmail.com
Joana Figueiredo²
jfigueiredo@ipatimup.pt
Raquel Seruca²
rsuruca@ipatimup.pt
J. Miguel Sanches¹
jmrs@ist.utl.pt

¹Institute for Systems and Robotics
Department of Bioengineering, Instituto Superior Técnico
Technical University of Lisbon, Portugal
²IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto
Portugal

Abstract

Structural mechanical properties of the tissues are dependent on the physical linkage between cells. E-Cadherin is the molecule that promotes this adhesion and changes in its molecular constitution are usually associated to loss of cell-cell adhesion which increase the invasion potential of cancer cells.

The E-Cadherin is usually mainly concentrated at the membrane where it promotes linkage between cells. When the E-Cadherin gene is mutated the distribution of the dysfunctional molecules is different and most of the time it is no longer near the membrane, were it plays the adhesion role.

In this work a computational tool is proposed to semi-automatically help in the segmentation of cells from microscopy images of fluorescence with tagged E-Cadherin and to compute an image of radial profiles of the molecule distribution from the center of the cell toward the membrane. The radial profiles are geometrically compensated, to cope with shape and size differences among cells, and a representative profile of the tissue is obtained for mutation detection purposes.

Examples with real microscopy images of fluorescence of epithelial cells of the stomach are presented to illustrate the method.¹

1 Introduction

Gastric cancer (GC) remains a major concern worldwide characterized by an inherent molecular heterogeneity and consequent divergent clinical biology. According to the World Health Organization (WHO) and the Lauren classifications, two main histological types of GC can be described displaying distinct clinicopathological features, the diffuse and intestinal subtypes..Diffuse type cancer is composed by non-cohesive cells (with or without signet ring cells) and is more commonly observed in younger patients [1]. Germline mutations of the E-cadherin gene(CDH1) are involved in the tumorigenesis of hereditary diffuse gastric cancer (HDGC).

It is well accepted that E-cadherin plays an important role as an invasion suppressor gene/protein, since its loss of expression, abnormal function, or both, leads to an increased ability of cells to invade neighbouring tissues, as verified in cancer.[3] To observe its distribution in the intra and inter cellular space an in vitro assay was developed. The distribution of normal E-Cadherin molecules in cells is mainly observed at the membrane, in order to accomplish its role in cell-cell adhesion. In this case, the image intensities are usually stronger at the membrane than at other locations. Additionally, the distribution of the molecules in normal cells at the cytoplasm is uniform with a characteristic distribution pattern (textural characterization). On the contrary, the E-cadherin in mutated cells follows a different distribution. Large concentrations in the cytoplasm or lack of E-cadherin signals in some locations are common features for almost all mutations that are associated with loss of cell-cell adhesion.

The objective of this work is the designing of a characterization algorithm for the E-Caderin distribution in the intracellular space specially near the membrane for mutation detection and discrimination purposes. The strategy to achieve this goal is composed by four main steps i) pre-processing and nuclei segmentation, ii) geometric compensation, intensity profile extraction of each nucleus and sample prototype profile computation, iii) features extraction and selection and iv) classification.

This paper addresses the semi-automatic cell segmentation from microscopy images of fluorescence [4], as shown in Fig. 1.a), and geometric compensation for shape variability (principal focus of this work).

The segmentation procedure is based on the Canny operator to detect the nuclei and the Hough Transform to find their geometric centroids. The method to extract the radial E-Cadherin distribution profile is also described. The differences between these profiles, located at the columns of the profile image, as shown in Fig. 1 (c), are mainly due to geometric differences among the cells, that are not perfect spheres. Therefore, for an accurate radial distribution characterization, these differences must be compensated.

The proposed method to make this compensation is based on a novel approach where the image profiles are modelled as a continuous field in $\mathbb{R}^2$ estimated from the original profiles adjusted along an iterative procedure. The algorithm, contrary to what usually happens, adjusts of observation locations driven by the minimization of an energy function.

![Image](320x412 to 546x512)

Figure 1: Geometric profile compensation, fluorescence Image (a), Nucleus segmentation (b) and radial profile (c).

2 Problem Formulation

The procedure described in this section intends to create a standard profile representative of the cells which compose the microscopy fluorescence image allowing a posterior classification of the images. To obtain this unique profile of E-cadherin a first pre-processing step is implemented and afterwards an iterative algorithm which aims to compensate the shape variability of the cells that are not related with mutations of E-Cadherin but prevent the computation of a radial prototype distribution of the molecule, representative of each mutation.

In the pre-processing step the radial profiles, centered at the centroid of the cell, for different angles, $\theta$, are organized in columns of a profile image, as shown in Fig. 1. In a pure radial distribution of the E-Cadherin, on a spherical cell, this profile image is angle invariant (across columns) presenting only information along the distance $\rho$.

In the first step of the iterative part of the algorithm, a finite dimensional continuous function, $f_k(x) = \sum_{i=0}^{N-1} c_k(i)\phi_i(x)$, is obtained to describe each angle profile (column). The column vector, $c_k = \{c_k(i)\}$, defining each of the $k$th column of the continuous map profile, is obtained minimizing an energy function

$$c_k = \arg\min E(x_k, y_k, c_k)$$

where $E(x_k, y_k, c_k) = E_x(x_k, y_k, c_k) + E_y(c_k)$. The data fidelity term, that pushes the solution toward the data, is $E_x(x_k, y_k, c_k) = \sum_j (f_k(x_k(j)) - y_k(j))^2$ and the prior term used to stabilize the iterative process and smooth the solution is $E_y(c_k) = \alpha \sum_{j=0}^{N-1} (c_k(j) - c_k(j-1))^2 - \alpha \| c_k \|^2$ where $\alpha$ is the prior hyper parameter. By using matrix notation this energy function is the following

$$E(x_k, y_k, c_k) = (\Phi c - y)^T (\Phi c - y) + \alpha (\theta c)^T (\theta c)$$
where $\Theta$ is a difference operator. The minimizer of (2) is computed by finding its stationary point, $\nabla_{\theta} E = 0$, that leads to the following solution,
\[
c^0 = (\Phi^T \Phi + \alpha \Theta)^{-1} \Phi^T y;
\]  
(3)
where $\Theta = \theta^T \theta$.

In the second step of the iterative process a prior term is added to the expression of the energy of the system with the purpose to force similarity between columns. The energy will become: $E(x_k, y_k, c_k) = E_p(x_k, y_k, c_k) + E_p(c_k, c_{k-1})$ where $\beta$ is the new prior hyper parameter and $E_p(c_k, c_{k-1}) = \beta \sum_{j} (c_k(j) - c_{k-1}(j))^2$. Using matrix notation this energy function becomes:
\[
E(x, y, z) = \sum_k (\Phi c - y)^T (\Phi c - y) + \alpha (\Theta c_k)^T (\Theta c_k) + \beta (c_k - c_{k-1})^T (c_k - c_{k-1})
\]  
(4)

The minimization of (4), results in:
\[
\Phi^T (\Psi c - y_k) + \alpha \Theta c_k + \beta (c_k - c_{k-1}) + \beta (c_k - c_{k-1}) = 0
\]  
(5)
which can be rewritten as:
\[
\sum x_k c_k - \beta C P_k = \Phi^T (x_k) y_k
\]  
(6)
where $\sum (\Phi^T \Phi + \alpha \Theta + 2 \beta I_k), \beta (c_k - c_{k-1}) = \beta C P_k, C$ is the matrix containing in the $k^{th}$ column the actual $c_k$ and $P_k$ the $k^{th}$ column of a shift matrix. Rearranging the terms in order to calculate the $c_k$ of each column we finally obtain:
\[
c_k = \sum x_k^{-1} (\beta C P_k + \Phi^T (x_k) y_k)
\]  
(7)

In the third and last step, the observations position will be adjusted after the computation of the vectors of coefficients, $c_k$. The energy function of this step is $E(x_k, y_k, c_k) = E_l(x_k, y_k, c_k) + E_p(x_k)$. where $E_p(x) = \gamma \sum p^2 (x_j - x_{j-1})^2$. In this case the minimization of this expression will be in order to $(x)$ doing $\nabla_{x} E = 0$ which will result in:
\[
2 (f_k(x_r) - y_r) f_k(x_r) + 2 (x_r - x_{r-1}) + 2 (x_r - x_{r+1}) = 0
\]  
(8)
Rearranging the terms, $x_k(i)$ can be calculated making:
\[
x_k(i) = 1/2 (Z_k(i) + \Psi_k(i)^T)
\]  
(9)
where $Z_k(i) = (f_k(x_k(i)) - y_k(i)) f_k(x_k(i))$, $x_k$ is the actual $k^{th}$ column of the matrix which collects the actual positions of the observations and $\Psi_k$ the $k^{th}$ column of a specific shift matrix.

3 Experimental Results

The proposed algorithms were implemented in Matlab$^\text{\textregistered}$ and synthetic and real data were used for illustrative purposes. In both cases first we estimated the coefficients using equation 3, in order to initialize $C^0$ and defined $\alpha$ unitary in this expression to remove noise. Afterward the iterative method of geometric compensation was implemented alternating equations 7 and 9. The final result of the geometric compensation can be seen in Figures 2 and 3.

3.1 Synthetic data

In this experiment a single 256 x 130 gray scale synthetic image was used containing an arc of cosine as shown in Fig.2.a. This aims at simulating distorted distribution of the E-Cadherin at the membrane. The result of the geometric compensation applied to this image is shown in Fig. 2.b.

3.2 Real data

For illustrative purposes a single radial profile created from a region of interest of a fluorescence microscopy image (see Fig.1.i) was used. The initial radial image size is 101x360 (size of the ROI* number of degrees of the radial profile). When computed we choose $N$ (number of vector basis) as equal to the number of lines ($N=101$). Some tests were proceed to estimate the better $\beta$ in this case in terms of stability and time of the program and obtained: $2 < \beta < 10$. With this parameters we proceeded 2000 iterations. The results are displayed in Fig.3.

4 Conclusions

In this paper a semi-automatic segmentation procedure of cells from microscopy images of fluorescence with tagged E-Cadherin, is described as well as the algorithm to extract the radial intensity profiles of them for distribution characterization purposes.

These profiles are affected by geometric differences between cells that are not perfect spheres, and are confound factors in the process of E-Cadherin mutation detection and discrimination.

In this paper an algorithm to compensate for these differences is described. The method is based on the estimation of a continuous field in $R^2$ based on moving observations driven by the minimization of an energy function containing specific priors that regularize the estimated field along the distance to the centroid of the cell.

In this paper we have presented an algorithm of geometric compensation which can be applied for instance in cases of biological variety in geometry. A future goal is to apply this strategy to more complex images and encounter the relation between $\gamma$ parameter and the noise which stabilizes the method.

References

Assisted Teleoperation of a Quadrotor using Active Perception

João Mendes
jmendes@isr.ist.utl.pt
Rodrigo Ventura
rodrigo.ventura@isr.ist.utl.pt

Institute for Systems and Robotics
Instituto Superior Técnico
Av. Rovisco Pais, 1
Lisbon, Portugal

Abstract

The teleoperation of UAVs often demands extensive training, since even well trained pilots are prone to mistakes, resulting frequently in collisions of the vehicle with obstacles. This paper presents a method to assist the teleoperation of a quadrotor using an obstacle avoidance approach. The target scenario is unknown, unstructured, and GPS-denied. A short-term rough map of the nearby environment is constructed using sonar sensors. This map is constructed using FastSLAM to allow tracking of the vehicle position with respect to the map. A danger classification method is then applied to choose the appropriate action for each particular, and potentially dangerous, situation. A simple active perception routine is used to orient one of the sensors to an unknown area, in case the UAV is ordered to move towards an unmapped area. Real world results are presented allowing a preliminary validation of the proposed methods.

1 Introduction

Quadrotors have several characteristics that make them well fit for situations where hovering is crucial, such as indoor usage. In particular, its high agility, small size and hover capacity make this type of vehicles useful for, as example, remote inspection of areas where ground vehicles are unable to reach. However, its teleoperation, particularly in confined environments, is far from trivial.

In this paper we address the problem of Assisting Teleoperation of a quadrotor. We define assisted teleoperation as the process of overriding the operator input, either by modulation, inhibition, or replacement with a different input. Our goal is to use sensor data to determine an appropriate assisted teleoperation in order to guarantee a safe flight.

The presented solution is based on a FastSLAM [5] approach using an occupancy grid map [1]. Since the purpose of this work is not a detailed map of the environment, the problem can be efficiently addressed by knowing the relative position of the quadrotor in relation to nearby obstacles. After knowing the vehicle’s position and map, a decision making method based on danger assessment is applied. This classifier overrides the user’s inputs if they compromise the quadcopter’s physical integrity in the near future. Overriding may range from simple velocity reduction to, in extreme cases, an evasive maneuver. As our aim is to ensure the vehicle’s safety at all times, an active perception methodology is applied to address map uncertainty whenever it is necessary. Unlike a purely reactive methodology, if the map is kept in memory it is possible to avoid crashes in sonar’s blind spots. The full architecture is presented in Figure 1.

Figure 1: Full architecture of the proposed approach

3D Simultaneous Localization and Mapping (SLAM) in Unmanned Aerial Vehicles (UAV) using lasers has been studied but typically including techniques, such as loop closure algorithms [4]. As for obstacle avoidance methodologies for UAVs, literature mostly addresses path re-planning topics [2]. This paper differs from the above in that we aim at a rough and low complexity map, and thus more time efficient.

2 Methodology

2.1 FastSLAM

Correct attitude is assumed to be maintained at all times by an onboard IMU. Since accurate attitude estimations can be provided by a commercial solution, the 6D problem (position and attitude) is reduced to a 3D problem (position only). The objective of SLAM is to estimate the robot’s position and the 3D map of the environment simultaneously and hereby solved by the FastSLAM approach proposed by Montermerlo et al. [3]. The conditional independence property of the SLAM problem allows for a factorization of the posterior and decompose SLAM into a path estimation problem and a mapping problem. This is solved by a combination of a Particle Filter with an occupancy grid mapping algorithm.

2.2 Decision Block

The inputs for the Decision Block are the position estimation and a map m. Each map cell i is classified in one of 3 states, c^i ∈ {F, U, O}, corresponding to Free, Unknown, Occupied. The classifier is based on thresholding the occupancy probabilities, according to:

\[ c^i = \begin{cases} 
F & \text{if } P(m_{occupied}^i) < 0.5 \\
U & \text{if } P(m_{occupied}^i) = 0.5 \\
O & \text{if } P(m_{occupied}^i) > 0.5 
\end{cases} \] (1)

All cells are initialized with \( P(m_{occupied}^i) = 0.5 \) thus classified as Unknown. After being mapped, they become classified as either Free or Occupied.

The global flow chart of the Decision Block is presented in Figure 2. When the user inputs a desired velocity to the quadcopter, the Active Perception block, Figure 3, validates if the commands will be applied to the vehicle without any constraint. To do so, the inputs are firstly used to compute the desired movement direction and the distance to the first non-Free cell. If the closest cell is Occupied, the inputs are subject to a confirmation that they do not compromise the vehicle’s safety in a near future and then applied to the vehicle. If the closest cell is Unknown and if the distance to it is larger than a certain threshold (TH) the algorithm checks whether there is any sonar aligned with the desired direction. If not, the algorithm will autonomously rotate the vehicle and then apply the same norm of velocity the user demanded. The algorithm is, therefore, able to avoid flying in unknown areas as the sonar is pointing towards the movement direction and still perform the order the user requested. By applying active perception it is possible to guarantee that the vehicle is not allowed to fly towards unknown areas.

Figure 2: Flow chart of the Decision Block.
An obstacle’s position is computed by a method designated here by volume check. The algorithm applies an extrema of a square centered on the quadcopter’s position, along the velocity vector $v$, as illustrated in Figure 4. The size of this square, $b$, encompasses the quadrotor volume while $a$ is the length of the volume in which obstacles are searched. This volume enables the algorithm to know which grid cells are in the near-future path of the vehicle, to determine the position of the closest occupied cell, and to compute the distance between that cell and the vehicle. It is then possible to predict how long it takes — if we maintain the current speed — to collide with it. This concept is known as Time To Collision (TTC) and is a crucial step in the classification of the danger levels.

The classifier block acts as a multiplexer by choosing an input, and forwarding it to the vehicle, given a certain TTC. The threat levels differ in the action performed. This action may be allowing the user to fully control the vehicle or to impose an override to the user’s inputs. This override ranges from ordering the vehicle to lower its velocity if the danger is low to, in extreme cases, an evasive maneuver.

3 Real world results

In this section, a preliminary evaluation to the architecture, combining the FastSLAM and the decision block in a real world experiment, is presented. The available quadrotor is equipped with an IMU capable of providing filtered values of yaw, pitch and roll. Four sonars were used and placed above each of the propellers. The sonars’ maximum range is three meters. All readings were recorded with the quadrotor’s motors off and the vehicle manually placed in the different poses throughout the tests.

The presented example considers 5 different stages illustrated in Figure 5. It is considered that the user is ordering a constant movement to the vehicle, given a certain TTC. The threat levels differ in the action performed. This action may be allowing the user to fully control the vehicle or to impose an override to the user’s inputs. This override ranges from ordering the vehicle to lower its velocity if the danger is low to, in extreme cases, an evasive maneuver.

Figure 3: Flow chart of the Active Perception block.

Figure 4: Graphical definition of the volume check. $V$ represents the velocity vector.

In Table 1, the first line shows the distance directly reported by sonars in each situation while the second one shows the distance between the position estimation and the first obstacle in its own map. The proposed solution keeps a constant track of the distance to the obstacle allowing the classifier to choose the best action in each situation. It is also noticeable that the algorithm, due to the acquired attitude, only updates the distance to the obstacle when the sonar is pointed to the direction of movement — position 1 and 4. In the last two rows it is possible to see the classification $C$ given to the closest non-free cell $i$ and the consequent action performed on the command inputs.

Table 1: Comparison between the distance both methods and consequent classifier decision.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\text{reactive}}$</td>
<td>3</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>$d_{\text{mapping}}$</td>
<td>3.25</td>
<td>2.34</td>
<td>1.38</td>
<td>2.28</td>
<td>1.11</td>
</tr>
<tr>
<td>$C^i$</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

| Classifier decision | User | User | Override | User | Override |

4 Conclusion

This paper presented an assisted teleoperation method for UAVs, based on short-term, rough mapping of the nearby environment. In particular, we targeted quadcopter vehicles operating in unstructured, unknown, GPS-denied environments. By knowing the environment, the algorithm is capable of overriding user inputs whenever the vehicle faces a potentially dangerous situation. Whenever confronted with an unknown area, the active perception routine forces the vehicle to point a sensor towards that area. The main objective was successfully achieved in simulation and partially evaluated in a real world scenario. As for future work, we are currently working in a full real world evaluation of the proposed method.

References


Using Phoneme Acoustic Models for Fast Word Spotting

Arlindo Veiga, Fernando Perdigão and Luís Sá
http://www.it.pt/person_detail_p.asp?ID=808

Department of Electrical and Computer Engineering, FCTUC,
University of Coimbra
Instituto de Telecomunicações - Polo de Coimbra, DEEC, Pole 2 of the University of Coimbra

Abstract
This paper presents a fast word spotting approach based on phoneme acoustic models. This approach simplifies the decoding algorithm of speech recognition process using a generic model and a model of the word to be detected. The ratio between the word model likelihood and the generic model likelihood is given to a classifier which was trained to detect relevant peaks values that indicate word occurrences. Experiments were performed on continuous speech utterances of the Portuguese TECNOVOZ database obtaining an accuracy of 94.8% against 1.0% false alarm. This result prompted the application of this technique in thousands of words or multi-word expressions to TV broadcast news programs which is expected to enhance our acoustic models.

1 Introduction
In speech processing, word spotting is a technique used to detect instances of a given word or sequence of words in a stream of speech signal. This technique has several practical applications, mainly in human-machine interactions using information retrieval in speech and multimedia documents.

The development in word spotting techniques has followed the evolution of techniques of automatic speech recognition. The first approaches used models based on Dynamic Time Warping (DTW) [1, 2] and progressed to Hidden Markov Models (HMM), [3]. Currently, word spotting approaches can be classified into three main classes [4]: i) based on large vocabulary continuous speech recognition system (LVCSR); ii) based on a phone lattice; and iii) based on acoustic models for fillers, background or anti-models. The first two approaches require a LVCSR, which is not a trivial resource for many languages (mainly due to the lack of good databases) resulting in systems with poor performance on non-clean signal and demanding of high computational requirements. The word spotting system based on the first approach examines the LVCSR output results leading to a performance highly dependent of the LVCSR word accuracy. The second approach uses a LVCSR as a pre-processing stage to create phone/lattices. The word spotting system uses theses lattices to search for word occurrences. The third approach requires fewer resources; however, it is difficult to build consistent models for fillers, background or anti-models for all words. Most of the word spotting system based on this latter approach uses the filler model proposed in [5, 6] that is composed internally by the states of all phone models arranged in parallel, as shown in Figure 1. The word spotting system presented in this paper follows the last approach with this filler model. The word model is made by concatenating the models of the corresponding phones sequence.

2 Decoding
A common word spotting system based on HMM acoustic models uses the Viterbi algorithm through a decoding network (task grammar) that allows the filler model to compete with the word model. The decoding process ends with a backtracking pass which determines where the word model “wins” the filler model on the speech utterance.

In this work we propose a new decoding process based only on model properties and on acoustic model's likelihoods. A score based on likelihood ratio and HMM information is proposed that allows decoding without the backtracking pass. In our case the filler model is not competing with the word model. Instead, it is used to compute the maximum likelihood of each observation frame given the phone models. The decoding process adds a feedback arc (dotted arc in Figure 1), allowing us to obtain the best phone sequence, thus with maximum likelihood (free phone loop), for the given speech signal and phone models. It is possible to define a measure of similarity between the output free phone sequence and the predefined word's phone sequence to detect the word occurrence, but this implies the backtracking pass and the definition of a similarity score. On the contrary, this work defines a similarity measure based only on the ratio between the model word likelihood and the filler model likelihood, avoiding the use of the phone sequence decoding. It assumes that the word model likelihood is similar to the filler model likelihood whenever the word occurs. Then, a fast decoder can be implemented using this similarity measure and a predefined threshold can be set to detect the spotted word occurrence. The similarity measure also uses the mean duration of the word models computed from the transition probability of states.

The proposed scheme to perform word spotting is shown in Figure 2. It performs a beam search step using the token-passing paradigm [7], to propagate likelihoods as well as frame-time information. It allows the winner token on the end state of the filler model to propagate to the initial state of the filler and word model. The time information is updated when the token propagates to the first state of the word model (only) and indicates a time stamp of a possible word beginning. This information overcomes the need of a sliding window strategy to obtain the word boundaries.

In each instant a winning token appears at the end of the two models. The token leaving the word model has the time stamp of a possible word beginning and the likelihood of the word hypothesis. If this likelihood is similar to the filler's, then there is a high probability of being in the presence of the spotted word. Otherwise, the likelihood is much lower than the filler’s, the word is probably not there. Although this likelihood ratio performs well in word detection, there are other restrictions that can be applied. For instance, the duration of the word segment hypothesis must be taken into consideration.

3 Similarity measure
We propose a similarity measure based on a likelihood ratio (or log-likelihood difference). Logarithm scale is used to avoid the precision issue on the likelihood computation. The log-likelihood difference between two sequences of N frames/states \( LL_{R_y} \), one given for word HMM and the other given for filler HMM, can be expressed as:

\[
\text{Score} = \frac{LL_{R_y}}{LL_{W_y}}
\]
\[ LLR_i = \sum_{n=1}^{N} \log \frac{L_{i,w}}{L_{i,f}} = \sum_{n=1}^{N} (\log L_{i,w} - \log L_{i,f}), \]  

where \( L_{i,w} \) and \( L_{i,f} \) are log-likelihoods of the \( i^{th} \) state of the HMM word sequence and filler HMM sequence respectively. This difference depends on the length of the segments \( (N) \). A reliable similarity score can be computed by normalizing the log-likelihood difference by the length of segments. So we define the first similarity score as:

\[ SS_1 = \frac{LLR_1}{N}. \]  

Besides the good performance of this similarity score, it does not take into account the average spotted word duration. Thus, we propose a second similarity measure based on the area defined by log-likelihood difference, as is suggested in Figure 3, normalized by the average word duration. This similarity score is computed as:

\[ SS_2 = \frac{1}{D_w} \sum_{n=1}^{N} (LLR_n - LLR_1), \]  

where \( D_w \) is the word duration estimated as the mean duration of the word model computed from self transition probability of the HMM states, \( D_w = \sum_{n=1}^{N} 1/(1 - a_n) \). This measure showed lower performance than \( SS_1 \), however we have found that the ratio between them,

\[ SS_1 = \frac{SS_1}{SS_2}, \]  

leads to better results than either of them individually. In all these three scores a peak picking is applied in order to obtain candidates of the spotting word.

4 Evaluation

The word spotting system has been evaluated on the Portuguese TECNOVOZ [8] speech database. A total of 22,627 utterances (around 31.5 hours) were used which corresponds to 208 prompts of generic sentences.

A set of acoustic models for 39 context free phones were trained with the HTK toolkit [9]. 37 models correspond to the phonemes of the Portuguese language; one model is used for silence and one model for short pause. The number of Gaussians per state (mixture PDF) was incremented gradually until reaching 96 Gaussian in the training stage.

To test the word spotting system, we select words with duration equal or greater than one second. We also use the utterances where the selected words do not occur in order to test the rejection thresholds. A total of 4288 utterances were selected making a total of 345 unique words. The current system only finds one word instance per utterance. The number of phones per word range from 3 to 19.

5 Results

It is common to use ROC (Receiver Operation Characteristcs) curves [10] or DET (Detection Error Trade-off) curves [11] to evaluate a binary decision system. We use a DET curve that is a plot of false rejection error (type I error) rate (FRR) against false acceptance error (type II error) rate (FAR), by varying a threshold. We can define a threshold for one operation point of word spotting system from the DET curve.

\[ \text{Equal error rate point} \]  

The equal error rate (where the FRR is equal to the FAR) is around 3.2% corresponding to a threshold of \(-0.1387 \) of \( SS_1 \). A real application usually requires FAR lower than FRR. With this measure, we can reach 1 % of FAR against 5 % of FRR. The performance of word spotting is highly dependent of the length of the word segment used. Longer segment has better performance. In this experiment we use segments with length equal or greater than one second. If a detected word segment has fifty per cent of overlap with the reference duration, it is considered a spotting word; otherwise it is ignored. There are also cases where no score peaks occur, which leads sometimes to a miss: this occurred 46 times, corresponding to 2% of miss rate. Some of these errors are due speech disfluencies (hesitations, filled pauses, repetitions) and utterances with low signal-to-noise ratio.

6 Conclusion

This paper proposes a fast and efficient decoding algorithm for word spotting system. The decoding uses log-likelihood difference and model information to compute a similarity measure. The measure presented achieved a good trade-off between the false alarm error and false rejection error and do not require a large vocabulary continuous speech recognition.

References

Word count variation in 1,092 human genomes

Sérgio Santos1
sergiomcm santos@ua.pt
João M.O.S. Rodrigues1
jmr@ua.pt
Vera Afreixo2
vera@ua.pt
Sara P. García1
spgarcia@ua.pt

1 Institute of Electronics and Telematics Engineering of Aveiro & Department of Electronics, Telecommunications and Informatics
University of Aveiro, Portugal
2 Center for Research and Development in Mathematics and Applications & Department of Mathematics
University of Aveiro, Portugal

Abstract

Computational methods to find patterns in DNA are usually based on the genome of a species [3], or on a comparative analysis of genomes of different species [7]. The one-species approach consists in comparing the frequency of words in the genome to the frequency of words predicted by a background model, like a Bernoulli or a Markov chain model. Here, we present a different one-species approach based on the genomes of multiple individuals. Instead of comparing the frequencies of words to a background model, we assess the word count variation between individuals. As a case-study, we use sequencing data for 1,092 individual human genomes made available by the 1,000 Genomes Project [1]. Considering words with 8 nucleotides, we analyse their CpG dinucleotide content and find the word count variation between individuals to be highly dependent upon this CpG content. This preliminary work showcases the potential of the multiple genomes approach.

1 Introduction

Strategies to find patterns in DNA sequences are generally based on single genome sequences. Such methods try to detect under or over represented words on the complete sequence, or on a subsequence derived from it, by comparing frequencies to a statistical model [3]. Under or over represented words are generally associated with some biological function, such as, transcription factor binding sites or gene regulation sites. The efficiency of these methods depends largely on the model and on the preprocessing of the input sequence either by defining a subset of sequences or by filtering certain regions of the genome, like low complexity regions [3]. Therefore, these methods have some limitations and tend to ignore words whose number of occurrences is well predicted by the statistical model. Strategies that compare genomes of different species are useful to evaluate conservation and selective pressure over time [6]. However, the solution they provide is limited by the size of the sample, that usually consists of a genome of each of the selected species. It is also common to use alignments on these type of methods, therefore the end result depends largely on the efficiency and limitations of the alignment algorithm. The number and the length of the sample sequences further restricts the applicability of these methods.

The method here presented is a different approach for the detection of DNA patterns. A sample of 1,092 individual human genomes [1] was used to try to overcome some limitations of existing methods. To demonstrate the potential of this multiple genomes approach, we analysed the variation of word frequencies in the 1,092 genomes, with a focus in words containing the dinucleotide CpG. This dinucleotide is under represented in mammalian genomes, being approximately 25% less frequent than GpC on the human genome. About 15% of the CpGs are located on CpG islands [6], continuously being generated and eliminated at a higher rate than other dinucleotides. This is a complex process that has been studied for some time but is not fully understood yet.

2 Methods

2.1 Dataset

Our dataset was generated using sequencing data from 1,092 individuals made available by the 1,000 Genomes Project [1]. To generate each individual human genome, we used the human reference genome assembly GRCh37.1 and a set of variant call format (VCF) files. The VCF files contain the alterations necessary to incorporate in the reference genome in order to obtain a different, individual genome. These alterations can be single nucleotide polymorphisms (SNPs), short insertions and deletions (indels) and large deletions. Each line of a VCF file contains the information for a single alteration. It identifies the chromosome, the sequence index and the type of alteration to be done. All possible alterations were incorporated. However, some overlapping alterations were found in the VCF files. In those cases, the alterations were merged, if possible; otherwise deletions had priority and were selected over SNPs and insertions (Table 1). This criterion was based on the heuristics of GATK tools and on the opinion of developers of the 1,000 Genomes Project. For this task, some computational tools from GATK [2] and vcf tools [4] were tested but in the end a C custom program was used because it performed significantly faster and more reliably.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) SNPs</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>(ii) Small insertions</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>(iii) Small deletions</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>1</td>
<td>ACG</td>
</tr>
<tr>
<td>(iv) Multi-base substitutions</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>1</td>
<td>AACGC</td>
</tr>
<tr>
<td>(v) Retroactive variants</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>(vi) Pseudo-concurrent variants</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
</tr>
<tr>
<td>(vii) Overlapping deletions</td>
<td></td>
</tr>
<tr>
<td>POS</td>
<td>REF</td>
</tr>
<tr>
<td>2</td>
<td>CTG</td>
</tr>
</tbody>
</table>

Table 1: Variants (i), (iii), and (iii) are processed similarly by our pipeline and version 1.4-37 of GATK [2]. From the variants associated to complex events (iv, v, vi and vii), cases (iv), (vi) and (vii) were not processed by GATK v.1.4-37.

2.2 Word Count Variation

Words of length 1 to 12 nucleotides were counted for each of the 1,092 individual human genomes. In order to determine which words vary more between the genomes, we calculated the coefficient of variation \(c_v\) for each word:

\[
c_v = \frac{s}{\bar{x}},
\]

where \(s\) is the standard deviation of the absolute frequency \(N\), or...
The histogram of the coefficient of variation ($c_v$) for words of length eight: (a) all words; (b) words with CpGs (red line) and words without CpGs (black line); (c) words with GpCs (red line) and words without GpCs (black line).

Table 2: Variation ratio for several classes of words: (i) words with at least one CpG compared to words with at least one GpC; (ii) words with at least one CpG and no GpC, compared to words with at least one GpC and no CpG; (iii) words with at least one CpG compared to words with no CpG.

<table>
<thead>
<tr>
<th>Word classes</th>
<th>$v_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) CpG/GpC</td>
<td>1.6</td>
</tr>
<tr>
<td>(ii) (CpG and noGpC)/(GpC and noCpG)</td>
<td>5.1</td>
</tr>
<tr>
<td>(iii) CpG/noGpC</td>
<td>5.5</td>
</tr>
</tbody>
</table>

The frequency of CpG words points to the lack of selective pressure to maintain the frequency of the word varies more between the genomes. Histograms are used to analyse the distribution of the $c_v$ values (Figure 1). To evaluate which, if any, dinucleotide has a determinant role on the distribution of the $c_v$, we plotted histograms for the words with and without a given dinucleotide (e.g., words with CpGs and words without CpGs). Those histograms were plotted for various lengths of words and for all dinucleotides but are here only shown for words of length 8 and for two dinucleotides (CpG and GpC) (Figure 1 b and c).

To infer how much words with a higher coefficient of variation vary with respect to other words, we define a variation ratio between two classes of words

$$v_r = \frac{\bar{c}_i}{\bar{c}_v},$$

where $\bar{c}_v$ is the mean of the coefficients of variation of the words in class $c_i$.

### 3 Results

The histogram of the coefficient of variation for all words with length 8 nucleotides show a clear separation of two classes of words (Figure 1 a). When plotting separately words with different dinucleotide composition, the histograms with the words containing the CpG dinucleotide versus words without the CpG dinucleotide (Figure 1 b) show a separation of two peaks observed in the previous histogram (Figure 1 a). The histograms of all the other dinucleotides did not show a separation of peaks and they were identical to the histograms of GpC dinucleotides (Figure 1 c). This implies that the dinucleotide CpG is responsible for the separation observed in the histograms and that words containing the CpG dinucleotide vary more between the genomes than words without CpGs. The same analysis was done for words of lengths 5 to 12 nucleotides and the results were similar. However, it should be noticed that the longer the word, the higher the overlapped area between the two peaks.

The variation ratio (Table 2) shows that the frequency of CpG words varies on average five times more than words without CpG (class iii). CpGs words when compared to GpC words (this last dinucleotide is generally used for the comparison with CpGs because both have the same nucleotide composition) show a variation ratio of approximately 1.6 (class i). This is because some words have both CpGs and GpCs, hence they contribute equally to the mean $c_v$ of each class. When these words (with both CpGs and GpCs) are removed from the ratio (class ii), its value becomes similar to the ratio of class (iii). This means that CpG words vary approximately the same with respect to both GpC and noCpG words.

### 4 Discussion

Our results show a clear separation between words with and without CpGs. As there are mutations leading to the generation and elimination of this dinucleotide, it means that it is not possible to explain these results by mutation rates only (though this seems to be a necessary condition), as high variation rates in words other than CpG words should also have been observed. This is in accordance with mutation rates described in the literature [5]. The frequency of CpGs is approximately the same between genomes and there is a selective pressure to maintain the total frequency of CpGs in the genome. The high coefficient of variation observed for CpG words points to the lack of selective pressure to maintain the frequency of words containing CpGs, i.e., there is no selective pressure to maintain the context of the dinucleotide CpG. This information was not possible to obtain with previous methods based on a single genome.

### 5 Conclusion and Future Work

Here, we demonstrate the potential of the multiple genomes approach and the type of information that can be extracted using a larger sample of a given population. This approach was used as a stand alone method but, in the future, we will study how to use it in complement to already established methods. Other word lengths will also be studied, aiming at finding patterns and correlations that may help further understand the human genome.

### 6 Acknowledgements

This work is supported by project grant FCOMP-01-0124-FEDER-010095 funded by FCT and co-funded by COMPETE, QREN, and FEDER. SPG acknowledges funding from the European Social Fund and the Portuguese Ministry of Education and Science.

### References


Evaluation of different incremental learning methods for video surveillance scenarios

Samaneh Khoushrou¹
skhoushrou@inescporto.pt
Jaime S. Cardoso¹
jaime.cardoso@inescporto.pt
Luís F. Teixeira²
luisft@fe.up.pt

¹ INESC TEC (formerly INESC Porto)
² Faculdade de Engenharia
   Universidade do Porto

Abstract
Motivated by the problem of segmenting and tracking objects in video streams, we investigate the application of incremental learning algorithms for evolving data to the development of robust and flexible video object tracking systems. Most existing learning, which are variations on the static learning schemes, can not cope with many real-life challenges, rendering the technology impractical. We propose to develop models that explicitly learn the number and identity of objects depicted in a given video stream. In this work, we investigate the application of three incremental algorithms for adaptive learning from evolving data streams.

1 Introduction
Much of the recent history of computer vision and visual data understanding in general has focused on very simplified settings in stationary environments, such as, fixed and known number of categories to be recognized or learnt, enough computational resources are available [1, 2]. Yet, if the ultimate goal of computational intelligence is to learn from large volumes of data that come from real applications, then the need for a general framework to learn from and adapt to a non-stationary environment can hardly be overstated.

We will focus on adaptive learning algorithms for evolving data applied to the development of robust and flexible video object tracking systems. In particular, we will tackle the problem of tracking objects in multiple video streams. Most existing models, which are variations on the static learning schemes, cannot cope with many real-life challenges, rendering the technology impractical. The main objective of our work is to contribute with a system capable of continuously extracting new information from streams of visual data, in non-stationary environments. Thus, there is the need for a general framework for learning from and adapting to a non-stationary environment. Given new data, such a framework would allow us to learn any novel content, reinforce existing knowledge that is still relevant, and forget what may no longer be relevant, only to be able to recall, if and when such information becomes relevant again in the future [3].

An ideal algorithm possesses incremental learning capabilities in a wild environment, if it meets the following criteria: a) Learn incrementally, b) Handle the concept drift, c) Accommodate new classes in Non-stationary environment, d) Bounded complexity and e) Ability to work with unlabelled data or partially labelled data and finally, to be more specific in video applications: f) Handle Multi-Dimensional(MD) data. In a non-stationary environment, where the number and identity of objects change, we need to deal with unlabelled data which makes use of clustering approaches at least in primary steps of process inevitable.

As material for our approach is the main trajectories of people in the scene, we are interested in clustering parallel data streams. This research field has been divided into two groups: clustering by example, and clustering by variable. In clustering by example data points from the same stream may have been assigned to different clusters. On the other hand, clustering by variable treats a stream as one unit and a stream with all its points is assigned to one cluster [4]. So, clustering by variable seems to be more efficient in terms of runtime and complexity. In such scenarios, we also may have access to partially labelled data or we may be able to define some classes after a while. To cover these situations, we also need an incremental classification approach which possesses all mentioned characteristics.

To the best of our knowledge, none of the existing clustering and classification approaches fulfil all our demands, however methods like ODAC [5], Correl-Clus [6], and Learn++.NC [3] can suit most of our system demands.

The paper is organised as follows. In the next section, two clustering approaches are reviewed and discussed. An incremental classification method is described in Section 3. A comparative study is presented in Section 4. Some primary results are presented in section 5. Section 6 will be devoted to future work. Finally, in Section 7 the paper is drawn to conclusion.

2 Clustering algorithms
Clustering refers to the process of grouping a collection of objects into groups or “clusters” such that objects within the same class are similar in a certain sense, and objects from different classes are dissimilar. Herein, two variable-based data stream clustering approaches are reviewed and their characteristics are discussed.

ODAC [5] is an example of variable based approaches, using tree models for on-line clustering. In this method each observation is processed only once and the system is incrementally updated. For updating, the system computes the dissimilarities between time series and considers the cluster’s diameter as the highest dissimilarity between two time series belonging to the same cluster. When a cluster’s diameter exceeds a threshold, the cluster is split. One important feature of this system is that the dissimilarity matrix is updated only for the cluster currently under test. So, the time complexity of iterations are constant given the number of examples. It may also decrease with every split occurrence. Moreover, the framework does not need a predefined number of target clusters. Furthermore, it provides a good performance on finding the correct number of clusters. However, there is no statistical confidence on the decision of assignment when the structure expands, which might split the variables. Besides, it can not handle dynamic number and length of high-dimensional data streams, which is more than likely situation in a non-stationary scenario.

In [6] a correlation-based framework, called Correl-Clus, is proposed. Unlike the widely-used Euclidean distance which only measures the discrepancy of the data values, this correlation-based distance considers two data streams with similar trends to be close to each other. Correl-Clus continuously receives raw data records from all the data streams at each time step. It keeps outputting the clusters for the most recent data streams of fixed length L. For every l time steps, it first computes a compressed representation of the data streams for the time segment [l-L+1, t], discards the raw data, and updates the compressed representation of the data streams for the target clustering time [t-L+1, t]. It then calls a correlation-based k-means algorithm to compute the clustering results. Since the number of clusters k may be changing, Correl-Clus also employs a new algorithm to dynamically adjust k in order to recognize the evolving behaviours of the data streams. This method is proposed to cluster one-dimensional data (1D), where the number and length of streams are fixed.

3 Classification algorithms
Classification is the association of patterns with same set of properties into classes. For on-line applications, system needs to deal with large number and various classes. So, using multiple classifiers could lead to better performance than could be obtained from a single one. Learn++NSE (non stationary environment), trains a new classifier for each new data it receives and combines these classifiers using a dynamically weighted majority voting. This approach allows the algorithm to recognize, and act according to changes in data distributions [3]. Although this method has shown the most promising performance, it has several shortcomings. For instance, if the ensemble is not pruned, the complexity of model will grow...
unbounded over time; on the other hand the typical solution of replacing
the oldest models with new one potentially discards important informa-
tion.

4 Comparative study

An objective comparison of the methods, already explained in previous
sections is presented in Table 1. It can be concluded, that none of the
methods has all the characteristics, so we can not apply them directly to
our non-stationary incremental scenario.

5 Experimental Results

The evaluation of already mentioned methods are made using synthetic
data sets created with specific definitions, described in following section.

5.1 Dataset

The data sets applied to evaluate clustering approaches were created using
a time series generator that produces n time series belonging to a prede-
fined number c of clusters with a noise constant β. Each cluster c has a
pivot time series p, and the remaining time series are created as p + λ where
λ \approx U(−βp, βp)

(1)

We created three data sets with ten variables each, observed along 7K
examples, with β = 0.3 and especially prepared to test different hypothesi-
son: one cluster(1C), two clusters(2C) and three clusters(3C). The structure of
3C was created as \{ {a_1, a_2, a_3}, {a_4, a_5, a_6}, {a_7, a_8, a_9, a_{10}} \}.

To evaluate the performance of Learn++.NC algorithm, Gaussian dataset
is applied. This dataset features multi class data, each drawn from a Gaus-
sian distributions. Each class experiences gradual but independent drift.

Figure 1: ODAC structure with 3c and 7k observations

A brief overview of three incremental learning algorithms was explained
and an objective comparison is presented. Hence, they showed good per-
formance in non-stationary scenarios, but none of these methods fulfil de-
mands for a multi camera video surveillance scenario. ODAC shows bet-
ter on-line clustering performance than Correl-Clus. To handle labelled
data in non stationary environments, Learn++.NSE is applied.

Acknowledgements

The authors would like to thank Fundação para a Ciência e a Tecnologia
(FCT) for the financial support of PhD grant SFRH/BD/80013/2011.

6 Future work

Since we are interested in scenarios where the data is not completely la-
belled, we will also research the clustering of data streams. These algo-
ithms are fundamental in our setting since, not knowing the classes a pri-
or, one is particularly interested in knowing if a certain object observed
at a certain time in a certain camera is the same object being observed in
a second camera at a different time. As an object may appear a short pe-
riod in scene, we will research algorithms working with dynamic number
and length of streams. Our methods should also handle multi-dimensional
data.

Ensemble-based approaches, herein Learn++.NC, suffer from an un-
controlled growth of complexity, will address this issue by researching
new methods to simplify an ensemble of classifiers by a new set with
lower cardinality. Intimately related with this issue is the concept of simi-
arity between two different classifiers.

7 Conclusions

A brief overview of three incremental learning algorithms was explained
and an objective comparison is presented. Hence, they showed good per-
formance in non-stationary scenarios, but none of these methods fulfil de-
mands for a multi camera video surveillance scenario. ODAC shows bet-
ter on-line clustering performance than Correl-Clus. To handle labelled
data in non stationary environments, Learn++.NSE is applied.

References

[1] L. Fu, “Incremental knowledge acquisition in supervised learning
[2] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A
in nonstationary environments,” IEEE Transactions on Neural Net-
[4] P. Chaovasit, “Clustering transient data streams by example and by
edge Discovery in Real Life Applications, vol. I-Tech Education and
On the compression of FASTQ quality-scores

Diogo Pratas
pratas@ua.pt
Armando J. Pinho
ap@ua.pt
Signal Processing Lab, IEETA / DETI,
University of Aveiro,
3810–193 Aveiro, Portugal

Abstract

The genomics sequencing centers and the scientific community are being flooded with data. A single sequencing machine can nowadays generate more data in one day than any existing machine could have produced throughout the entire year of 2005. The pressure to find efficient compression algorithms for next-generation sequencing (NGS) data is being felt worldwide.

The NGS most used format, FASTQ, contains three channels of information, namely headers, DNA bases and quality-scores (QS). Together with the DNA bases, the QS are the most demanding components in terms of storage requirements and compressibility. In this paper, we analyze the QS using several compression techniques aiming to find the best compression model that describes this type of data.

1 Introduction

The scientific community is currently facing a shift in concern that was probably not foreseen until recently: Genomic data are being produced at a pace that exceeds the growth of the media capacity to store them. Nowadays, it is common to find genomics sequencing projects having a larger fraction of the budget allocated to the computational infrastructure (including the storage component), than to the biological part.

Modern sequencing instruments are able to generate at least hundreds of millions of short reads of genomic data and to store them in files with a specific format, such as the most used FASTQ [4] format. The FASTQ format, originally developed at the Wellcome Trust Sanger Institute, is a text-based format for storing biological sequences (DNA bases) and its corresponding QS, where both are identified with headers and encoded with ASCII.

Several compression algorithms have been proposed to date. Tembe et al. proposed G-SQZ [12] and DSRC [5] was proposed by Deorowicz et al. Both methods split the data into separate channels, namely headers, sequences and QS, and compress them individually using the Lempel-Zip algorithm and Huffman coding.

Kozenitis et al. and Hsi-Yang Fritz et al. proposed, respectively, SlimGene [9] and mzzip [6]. These are reference-based compression methods that exploit the redundant nature of the data by aligning reads to a known reference genome sequence and storing genomic positions instead of nucleotide sequences. Wan et al. proposed Q-scores [13], a method based on lossless and lossy transformations for the compression of FASTQ QS. Very recently, Jones et al. proposed Quip [7], a lossless compressor based on the Markov property, reference-based compression (for the DNA bases) and arithmetic coding, with fast execution and low memory usage. However, these algorithms seem to have difficulties when compressing the QS. Therefore, this work aims to study the nature of the QS, using compression techniques, in order to serve as a starting point for a future state-of-the-art compression algorithm.

2 Methods and Results

Compression can be seen as a two stage process: modelling and coding. Modelling is the process of studying the characteristics of the data source to estimate the probability distribution of data. Coding involves the transformation of data into a sequence of bits according to the probability distribution acquired in the modelling process. Therefore, compression of QS data is a problem that is tightly related with QS information modelling. In order to study the model that better describes the QS, we have compressed ten sequences from the 1000 Genomes Project.

The results presented in Table 1 show better ratios for the compressors based on finite-context modelling (FCM), comparing with BZIP2 (based on the Burrows-Wheeler transform [3] and LZ77 algorithms) and GZIP (based on the DEFLATE algorithm). Moreover, for the most of the files, an order-3 seems to produce the best compression ratios.

Thereafter, we have implemented a finite-context model [1] using a sparse template (c135s2), as can be seen in the second layer of Fig. 1. This implementation improves the compression ratios, if compared with a regular template (presented in the first layer of Fig 1).

Since in a previous work [11] we have explored a three-state model for compressing the coding regions of DNA sequences, characterized by a periodicity of three (codons have three nucleotides), and gathering the information that the QS at a given position is highly correlated with the score at the preceding position [9], we have explored a two-state finite-context model.

A two-state finite-context model is causal, and the decoder is able to reproduce identical probability estimates without side information. Figure 2 shows the two-state finite-context model using a sparse template with context order of three, exploiting the two-symbol periodicity of the QS.

As it can be seen in Table 1, the two-state finite-context model using an order-3 sparse template (c135s2), on average, attains the best compression ratio.

Most of the recent NGS compression algorithms explore similarities between DNA bases (for instance in reference-based compression), achieving compression ratios better than those that do not use it. Therefore, our intention is to study the QS with the aim of discovering if there are similarities between different QS files.

In order to study the existence (or nonexistence) of similarities, we used the Normalized Compression Distance (NCD). Li et al. [10], aligning ideas from Kolomogorov [8] and based on an information distance proposed by Bennett et al. [2], defined this practical analog based on
Table 1: Rates (bits per base) for GZIP, BZIP2 and FCM-C compressors over ten sequences. The number of symbols of the alphabet is represented by $n$. In order to find the best compression model that describes this type of data, which appears to be two-state finite-context modelling using sparse context order of three. Moreover, we have seen that a reference-based comparison is not ideal to this type of information, since there are not similarities between QS files. Several hypotheses for testing on the other hand, if knowing $A$ and $B$, respectively compress the concatenation of $A$ and $B$ and NCD denotes the number of bits needed to compress the concatenation of $A$ and $B$. Notice that, ideally, NCD($A,A$) = 0 (this implies C($AA$) = C($A$)). However, practical compression algorithms usually yield NCD($A,A$) > 0. On the other hand, if knowing $A$ does not reduce the size of B, then we have C($AB$) = C($A$) + C($B$) and NCD($A,B$) = 1. In conclusion, smaller values of NCD($A,B$) indicate more similar sequences. According to Table 2, the NCD values are all greater or equal to one, with the exception of the diagonals (C($AA$) scenario) and the relation of Id10 with Id4 in both directions (still remains close to one). Therefore, overall, there are not similarities between QS files, leading to infer that a reference-based compression should not be the best implementation for this type of information.

3 Conclusions

In this paper, we have analyzed the QS, using several compression techniques, in order to find the best compression model that describes this type of data, which appears to be two-state finite-context modelling using sparse context order of three. Moreover, we have seen that a reference-based comparison is not ideal to this type of information, since there are not similarities between QS files. Several hypotheses for testing multiple models will be considered future work.

Funding

Supported by the European Fund for Regional Development (FEDER) through the Operational Program Competitiveness Factors (COMPETE) and by the Portuguese Foundation for Science and Technology (FCT) in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011).

References

A Two-stage Mass Segmentation Method for Breast Ultrasound Images

Rafael Rodrigues
jrafael.ubi@gmail.com

António M.G. Pinheiro
pinheiro@ubi.pt

Abstract
Breast ultrasound images offer several attractive properties which make them suitable and an interesting tool for breast cancer detection. However, due to their intrinsic high noise rate and low contrast properties, mass detection and segmentation becomes a challenging task. In this paper, a semi-automated two-stage breast mass segmentation method is proposed. Initially, ultrasound images are segmented using Support Vector Machines or Discriminant Analysis with a multi-resolution pixel descriptor extracted using non-linear diffusion, band-pass filtering and scale-space curvature. A set of heuristic rules complements the initial segmentation task, selecting the ROI in a fully automated manner. In the second segmentation stage, two different methods are used to attempt a refined segmentation of the area retrieved in the first stage. The first uses an AdaBoost algorithm, using curvature measures and non-linear diffusion of the original image at lower scales, and in the second active contours are applied to improve the spatial resolution of the ROI.

1 Introduction
Breast cancer is one of the major mortality causes among women, with special incidence in developed countries. Early detection and diagnosis of masses and nodules is crucial to reduce the mortality rate [2, 10]. Imaging techniques are an important aid for breast cancer diagnosis. Several different techniques are commonly used by radiologists including ultrasound images, mammography and magnetic resonance. Ultrasound imaging is a harmless, non-invasive and cost effective diagnosis method that offers real-time diagnosis capabilities. For its attractive properties it has been emphasized as a valuable tool for early breast cancer detection and diagnosis [2]. Due to the image acquisition process, ultrasound images typically contain a large amount of high-frequency noise rates and low contrast properties. In addition, this type of image commonly present shadows and other artefacts that result in a difficult analysis of the exam, even for experienced radiologists [2]. Thus, computer-aided diagnosis becomes an important help factor for radiologists. Several image segmentation methods have been proposed to respond to this problem [2, 3, 4, 10]. Nevertheless, many segmentation methods require manual selection of the region of interest (ROI) and typical ultrasound image characteristics make the segmentation of masses a difficult task. The proposed framework implements a two-step automated method, depending the selection of the final result on human decision. The first stage uses Support Vector Machines (SVM) or Discriminant Analysis (DA) to obtain an initial segmentation. A second approach based on the initial segmentation uses AdaBoost classifiers or active contour segmentation to improve spatial resolution.

2 Methods
2.1 Initial segmentation
The ultrasound images were initially segmented with the method described in [7], applying SVM or DA with several multi-resolution features. The features included in the pixel descriptor were the non-linear diffusion [5] of the original image after 60 iterations using the Charbonnier conductivity equation, with $\lambda = 1/8$ and using an adaptive $K$ set to match 90% of the integral of the gradient histogram. The features were classified with a SVM classifier [6] using the Linear, Quadratic, Polynomial, Radial Basis Function and the Multilayer Perceptron kernels and a DA classifier using the types Linear, Quadratic, Diagonal Linear, Diagonal Quadratic and Mahalanobis Distance. The SVM and DA training dataset consisted on the aforementioned descriptors’ information, retrieved from 654 pixels randomly selected from 2 ultrasound images. The respective binary segmentation masks have been manually segmented by an oncologist. The training dataset was formed with 323 positive pixels and 331 negative pixels, retrieved from two strips around the mask edge. This segmentation step was performed in a set of 20 ultrasound images containing masses. Considering the results of the initial segmentation task, heuristic methods were applied to those results as an attempt to select the ROI and eliminate missclassified regions. Objects connected to the image borders were suppressed and the output of this operation was subjected to a binary filling operation, closing holes inside regions. In order to retrieve a ROI for subsequent segmentation tasks it is desirable to select only one region among all regions classified as foreground. Thus, the largest area object was selected which is used as a starting point for the refining segmentation stage.

2.2 Refined ROI segmentation
The second stage of the proposed segmentation method attempts to improve the spatial resolution of the initial segmentation. For that purpose, two distinct methods were applied. In both methods, the results from the largest area selection performed in the first stage are used as starting points, indicating the ROI location. Considering that several SVM and DA kernels were tested in the initial segmentation of the images, yielding a set of 10 results for each image, a simple selection criteria was imposed to simplify the analysis of the refined segmentation performance. Only results with better accuracy were used.

2.2.1 Segmentation using AdaBoost
One of the methods applied to refined segmentation of the ROI uses an AdaBoost classifier [6] on a dilated region around the initial segmented area. The AdaBoost algorithm relies on a simple weak classifier, that establishes a threshold for each feature or data dimension individually, which divides data samples in two classes. This classifier is called iteratively, adapting the threshold in each data dimension to minimize the classification error. The training phase yields a model that stores the sequence of data dimensions and thresholds that minimize the training error. The pixel descriptor used for this segmentation step was constructed with features similar to some of those applied to the classifiers in the initial stage. However, different parametrization was used allowing the improvement of the spatial resolution of the output. Moreover, focusing on smaller areas around the masses directly reduces the presence of noise and other objects that competed directly with the ROI on the initial segmentation. Thus, lower scale curvature measures and non-linear diffusion outputs were chosen, based on their better spatial resolution when compared to the scales used previously. The AdaBoost training samples were retrieved based on the true segmentation mask of the training image. The negative training dataset retrieved 50% of the pixels located in a strip defined around the segmentation mask. The positive training dataset was formed using 20% of true positive mask pixels. The training dataset contained 2429 points, from which 1392 were positive samples and 1037 were negative. The classifier was trained and applied to the images with 200 iterations. The features included in the pixel descriptor were scale-space curvature measures at scales $t = 20$ and $t = 30$ and non-linear diffusion of the original image after 30 iterations using the Charbonnier conductivity equation, with $\lambda = 1/8$ and adaptive $K$ set to match 90% of the integral of the gradient histogram. In the clas-
sifying phase, the initial segmentation masks were dilated with a circular shaped structuring element with radius 8. This value was defined after experimentation, to maintain a trade-off between the selection of the biggest possible area without losing focus on the ROI. Only the pixels inside the dilated region were considered as inputs to the AdaBoost classifier, while pixels outside this area were automatically considered as background.

### 2.2.2 Segmentation using Active Contours

The other method applied for segmentation refining of the masses relies on active contours without edges. The algorithm implements the method described in [1]. Similarly to the AdaBoost methods, results from the SVM and DA segmentation were selected using the maximum accuracy criteria. The initial ROI segmentation was used to define the initial contour for the algorithm.

Active contours were applied with 200 iterations. The output of the algorithm was submitted to the selection of the largest area object, in order to eliminate any objects resulting from the defragmentation of the main contour, to produce the final segmentation.

![Manual segmentation](image1.png)  ![Initial SVM or DA segmentation](image2.png)

![AdaBoost segmentation](image3.png)  ![Active contour segmentation](image4.png)

Figure 1: Example of segmentation outputs.

### 3 Results

In table 1 are presented the overall segmentation performance measures. The somewhat low recall rates are due to two major factors. Firstly, the largest area criteria used to select the ROI directly reduces the efficiency of the algorithm in cases with multiple ROI. In this database, some of the images present two masses. Although the refined segmentation methods are applicable to more than one region simultaneously, further application results would be limited *a priori* (figure 1(a), 1(b)). Another factor, directly related to the properties of the initial segmentation algorithm, is the loss of spatial resolution resulting from high non-linear diffusion and curvature scales. Nevertheless, using lower scales on the entire image could result in the loss of ROI detection and increase the segmentation noise. For this reason, the segmentation refining methods are applied. Furthermore, human segmentation masks used to evaluate segmentation performance tend to define larger contours, due to poor edge definition of the images, which also has an influence on recall rates.

<table>
<thead>
<tr>
<th>Method</th>
<th>Initial</th>
<th>AdaBoost</th>
<th>Active Contour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.9727</td>
<td>0.9758</td>
<td>0.9770</td>
</tr>
<tr>
<td>Recall</td>
<td>0.6674</td>
<td>0.7646</td>
<td>0.7540</td>
</tr>
<tr>
<td>Precision</td>
<td>0.8830</td>
<td>0.8726</td>
<td>0.8751</td>
</tr>
</tbody>
</table>

Table 1: Segmentation performance measures.

Both proposed pathways achieve promising segmentation results, however, with some drawbacks, mainly regarding the overall recall rates. These could influence the clinical decision, leading to inaccurate diagnosis in terms of tumor size and shape. Although the refinement of the segmentation boosts the final results, the improvement of the training samples in the first stage (SVM or DA classification) or the introduction of new features may help to further reduce this limitation of the proposed methods.

### References


A PCA Approach to ECG-based Biometrics

Marta S. Santos
Ana L. N. Fred and Hugo Silva
André Lourenço
Instituto Superior Técnico
Instituto de Telecomunicações, Instituto Superior Técnico
Instituto de Telecomunicações, I. Superior de Engenharia de Lisboa

Abstract
Recent studies have revealed the potential of Electrocardiographic signals as a biometric modality for user identification. However, improvements can still be achieved concerning feature extraction and classification performance. In this paper we propose two identification methods based on Principal Components Analysis, and using a K-nearest-neighbour classifier. These methods are compared and, ultimately, a 0% error rate is shown possible.

1 Introduction
The heart is a myogenic muscle, which means that it is capable of generating its own electrical stimuli needed for contraction. The potential difference originated by that stimuli can be recorded and it is known as electrocardiographic (ECG) signal. Human hearts are similar from individual to individual and produce comparable electrical records. However, there are no exact copies; there is anatomical / physiological diversity in different individual’s hearts, and therefore an ECG signal differentiation between individuals is to be expected. Through pattern recognition techniques, this variability can be detected, allowing the recognition of different individuals based on their ECG signals.

A user identification system comprises two main phases: enrolment and identification [9]. On the enrolment phase, biometric templates are created and labelled. It requires signal acquisition and processing, followed by the storing of the labelled templates. On the identification phase, new samples of the biometric modality are acquired and processed. The new data is then compared with the stored one. The similarity between the new data and its best matching template is determined and compared with a threshold. The systems’ answer will then be “you are individual X” or “you are not enrolled”.

2 Pre-processing

We propose a three step pre-processing procedure, detailed next: filtering, segmentation and outlier removal.

![Figure 1: ECG pre-processing (framed in blue) and context](Image)

2.1 Filtering
The ECG signal acquired at the finger level has a fairly higher amount of noise than, for example, a chest ECG acquisition, and it can be affected by several noise sources such as motion artifacts, powerline interference and electromyographic noise. We designed a FIR band-pass filter (order 301, using a Hamming window), with 5-20Hz cutoff frequencies, to limit the bandwidth of the raw input signals, and thus reduce, or even eliminate, some of these artifacts.

2.2 Segmentation
We built on the work of Engelse and Zeelenberg [10] for offline QRS detection, adapting their algorithm for real-time QRS detection. A detailed description of the algorithm and comparison with offline approaches can be found in [11].

2.3 Outlier Removal

Making use of the R peaks detected, each heartbeat template is defined as the signal within a 0.6 seconds window going from 0.2 seconds before the R-peak, until 0.4 seconds after the peak. Since the sampling rate was of 1000 Hz, one considers each template as a sequence of 600 points, with the R-peak centered at point 200.

3 Template Creation and Decision Process

We propose two methods for feature extraction and template creation based on PCA, namely: Overall Population Eigen-Heartbeat identification and Individualized Eigen-Heartbeat identification. From the processed data, AS train segments are randomly chosen; the rest of the segments are used for testing. This way we guarantee that data used for training is not simultaneously used for testing.

3.1 Overall Population Eigen-Heartbeat (OEigHB)

In this method, the database will be formed by the mean heartbeat (HB) signal for the overall population, and corresponding eigenvectors obtained by PCA. Each enrolled ID is represented by the coefficients obtained from the projections of their heartbeats into the population’s eigenvectors. Those coefficients will be the templates that will represent each user in the database.

In the decision process, pictured in figure 2, the newly presented HBs are projected, one-by-one, into the population’s eigenvectors. The coefficients thus obtained are matched against the users’ templates, using the Euclidean distance, and the K-nearest-neighbor (KNN) templates are identified. Decision on the user ID uses majority voting.

![Figure 2: Overall Population Eigen-Heartbeat identification.](Image)

3.2 Individualized Eigen-Heartbeat (IEigHB)

In this approach, each user’s template is kept as the mean ECG segment, eigenvectors and coefficients corresponding to the projection of enrollment HBs into the individual eigen-heartbeats.
In the decision process, pictured in figure 3, the new HBs are projected into every enrolled user’s eigenvectors, and compared with that user’s coefficients. For matching and decision, similarly to the OEigHB approach, Euclidean distance, KNN and majority voting are used.

### 4 Experimental Setup and Results

#### 4.1 Acquisition Setup and Datasets

The ECG datasets used in our experiments were provided by the Check Your Biosignals Here initiative [12]. It comprises a set of one session acquisition of different electrophysiological signals, including the ECG, from a population of 65 individuals, while they were listening to the explanation of experiment’s goal and consequently visualization of a video sequence triggering different emotions, leading to a variable acquisition time.

#### 4.2 Number of Segments for Enrolment (NS)

From the previously processed dataset, NS randomly chosen segments were used as training data (enrolment templates), being the remaining ones used as test data. The NS parameter is related with the time required for enrolment. The following plots show the average and standard deviation of the error rates obtained over 25 runs of this procedure, using a 1-Nearest-Neighbor classifier.

Two values of NS were tested (20 and 30) for three data setups: without outlier removal; one-pass outlier removal; and two-step outlier removal. As shown in Figure 4, the higher the NS, the lower the error probability is. Due to the intra subject HB variability, a higher enrolment time will better capture that variability and is more likely to correctly identify the individuals. The OEigHB seems to be less sensitive to the various configurations, having lower ranges on the error rates than the IEigHB approach.

#### 4.3 Number of Segments for Identification (k)

The number of segments required for accessing the system relates with the acquisition time in the post-enrolment phase. It is the number of heartbeats, k, that the user needs to provide to the system in order to be identified. Again, due to intra individual HB variability, a higher number of segments will more likely match the template data, resulting in a lower error probability (see figure 4). The identification time does not need to be as large as the enrolment time. Figure 4 shows that for values of k between 9 and 15 the error variation is small, and for larger values, the error can be as low as 0%.

#### 4.4 Sensitivity to Outliers

Figure 4 shows that there is a clear benefit in putting the data through an outlier removal procedure since the error rate decreases considerably when comparing DataP with DataF. However, the gain is not so noticeable when comparing DataF with DataFF. Both methods are capable of achieving an error below 5% with no outlier removal if the NS and k are big enough. However, with DataFF both methods can achieve an error probability close to zero, and with a small variance.

#### 5 Conclusions

We proposed and compared two methods for ECG-based user identification based on PCA and KNN classification. Different design parameters were tested, and results showed that a 0% error probability is achievable. Template destabilization may occur through time and emotional variation lowering the accuracies obtained. This subject has not been thoroughly studied in the context of biometry; however some works are already trying to compensate this instability [13]. Ongoing work includes a further enlargement of the population set, and extending this study to situations of multiple acquisitions at distinct time instants.

### References


[12] Check Your Biosignals Here: Experiments on Affective Computing

Use of Artificial Neural Networks for Identifying Forces on a Bulldozer Ripper

Miguel Curinha Samarra
Luis Manuel Roseiro
Maria Augusta Neto
Tony Da Silva Botelho

Abstract
This paper presents a methodology for the monitoring of forces acting on a bulldozer ripper (claw-like device found at the rear of the bulldozer) to avoid overloading and fracture in service conditions. The methodology that is presented is tested using experimental data collected in laboratory and neural network technique as a tool for identification. The promising results obtained are presented and discussed.

1 Introduction
The ripper is a mechanical component attached to the rear of a trawl machine or bulldozer, and it main function is to penetrate and tear soil, particularly where it is not possible to use small size machines. It is usually used for doing works of high stress and fast jobs, such as in quarries for stone revolver, in mines for ground penetration or in works of levelling and uniformity of land. The ripper integrates the rear of the machine, operating as a complement to the main unit (paddle), which is hydraulically driven. Figure 1 shows a machine of this type levelling the ground in the construction of a motorway.

When ripper fracture occurs, given the size and costs associated with the ripper replacement, companies choose to weld the fractured parts instead of the purchase of a new ripper component. However, even under this procedure, the costs are high. In fact, the only possible way to avoid undesirable downtime of the bulldozer is to have an available ripper for replacement. When the welding process becomes unfeasible, another option is to produce a new ripper. Nevertheless, the ripper fracture is a situation that occurs frequently, and we strongly believe that the monitoring of forces during service conditions could allow the machine operator to decide for the stop of the ripper work instead of allowing that ripper force reaches threshold values for the material, and this would be a decisive way to reduce costs associated with their use. However, given the randomness and variety of soils to identify, this is not a linear process and is not easy to implement. In fact, when the ripper is inserted in the soil, it can encounter obstacles (usually stones) with varying dimensions that could be attacked by the ripper in different positions. This paper presents the first step of a methodology for early identification of forces acting on the ripper under working conditions. The methodology makes use of an artificial neural network with the data collected from electrical resistance strain gage. The methodology is tested with data collected in a laboratory environment, and the results obtained are presented and discussed.

2 Experimental Methodology
In order to obtain the necessary data for the implementation of the methodology, a mechanical system was developed and built. The main idea was to implement the set ripper-support connection as similar as possible to the working conditions. Figure 2 shows the geometric model of the ripper-support assembly, which was CNC machined to a scale of 1:10. The ripper was produced in 7022 aluminium alloy material and the supporting structure of the machine in AISI 4340.

The ripper was instrumented with 3 unidirectional strain gages, from Vishay (Micro Measurements Group) reference EA-13-250BG-120, grids made with constantan alloy, with auto-compensation on temperature for aluminium and a resistance of 120 ohms. The nominal gain factor of calibration certificate is 2.1 ± 0.5%. The position of the gages in the ripper was defined based on the information from a developed finite element model made with SolidWorks simulation and Abaqus CAE. Figure 3 shows the ripper instrumented and the position of the 3 unidirectional strain gages.

The monitoring of the ripper was assured by strain gage measurements. The ripper was instrumented with 3 unidirectional strain gages, from Vishay (Micro Measurements Group) reference EA-13-250BG-120, grids made with constantan alloy, with auto-compensation on temperature for aluminium and a resistance of 120 ohms. The nominal gain factor of calibration certificate is 2.1 ± 0.5%. The position of the gages in the ripper was defined based on the information from a developed finite element model made with SolidWorks simulation and Abaqus CAE. Figure 4 shows the ripper instrumented and the position of the 3 unidirectional strain gages.

Figure 1: Bulldozer working

Figure 2: Geometric model (a) and scale model (b) of the ripper-support assembly

Figure 3: Instrumentation of the ripper with strain gages
quarter-broad configuration and the load cell is connected in full-broad. Figure 4 shows the mechanical system and the experimental setup.

Figure 4: Experimental setup

3 Neural Network Identification Methodology

Among the several architectures used in practice, feedforward type neural networks, shown in Figure 5, have been considered suitable for the identification problem of the signature analysis. A feedforward neural network consists of several layers; each one with some processing elements, called neurons, linked to each other by weights, which determine the nature and the strength of the connection between the neurons. The number of nodes considered in the input and output layers depend on the specifications of the problem. The number of hidden layers, the number of neurons in each hidden layer as well as the activation function type for each neuron is selected according to the experience and some convergence criterions. The application of artificial neural network consists of two stages, namely training and testing. During the training stage an input-to-output mapping, using the available sample data, is presented to the network. The network evaluates its own output based on the presented input and compares this value with the target (presented) output. The actual output error (the sum of squares error function in this study) is used to adjust the node weights so that the error can be reduced. The learning stage stops once a cross validation pre-set error threshold is reached, and then the node weights are frozen at this point. During the testing stage, data that have not been presented to the network in the learning stage are provided as input and the corresponding output is calculated using the fixed node weights.

3.1 Developed Neural Networks

The main objective of the study is to estimate the horizontal force and its position based on the deformation collected from three strain gages placed according to the results obtained from a developed finite element model. The data training set should be representative of a broad class of possible input-output pairs. Thus, it is important that the network successfully generalize for the entire population that has learned. For a network to generalize successfully, intern parameters of the net (weights and bias) must be lower than training patterns [2]. In this study we consider a feedforward type network with four layers. A total of 3200 data sets were acquired, with the range of variation from 10 N to 500 N for the force and 1.23 µm to 252 µm for the deformations. The angle of force application ranged 0° - 90° and eight discrete possibilities [0, 10, 20, 30, 45, 60, 70, 90] were considered in data acquisition. The programming (definition and training) of neural networks considered was made in Matlab. The number of neurons in the layers of input and output have direct dependence of data input and output type desired. The inputs are the three deformations collected from the strain gages and the outputs are the force value and its relative position. It should be explained that the direct problem is: when applying a force to the ripper with some relative position, the strain gages indicate the deformation on the point where they are placed. In the inner layers, the choice of the number of neurons is based on experimentation with the minimum number that allows learning desired. Figure 5 illustrates schematically the idea and the network considered. The training of the neural network have been performed with a second order type algorithm, the Levenberg Marquardt [3] and various combinations of number of neurons in the inner layers were tested. From the acquired data, 2240 (70%) were used to train the networks, 480 (15%) for the cross validation process and 480 (15%) to test the networks. The selection of the data for training, validation and test were random for each network tested.

Figure 5: Feed forward neural network

3.2 Obtained Results

A total of 320 different neural networks have been tested. The number of hidden neurons varies from 8 to 15 as well as the type of activation function. Figure 6 shows the evolution of the identification errors with the configuration. The best results were achieved for 10 neurons in the first hidden layer, 14 neurons in the second hidden layer and the hyperbolic tangent sigmoid transfer function for all neurons. This led to the 3-10-14-2 network. Table 1 shows the relative errors of identification for the 5 best networks obtained in this configuration. The mean identification error was 4.8% for the force applied and 12.0% for the angle of the applied force.

Table 1: Mean identification errors in the five best networks

<table>
<thead>
<tr>
<th>Neural Network</th>
<th>Identification Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force Angle</td>
<td>Force Angle</td>
</tr>
<tr>
<td>1</td>
<td>4.4 12.1</td>
</tr>
<tr>
<td>2</td>
<td>5.8 10.9</td>
</tr>
<tr>
<td>3</td>
<td>4.9 11.4</td>
</tr>
<tr>
<td>4</td>
<td>4.4 11.7</td>
</tr>
<tr>
<td>5</td>
<td>4.5 14.0</td>
</tr>
<tr>
<td>Mean Error</td>
<td>4.8 12.0</td>
</tr>
</tbody>
</table>

Figure 6: Evolution of the identification with network configuration

4 Conclusions

In this study, a CNC machined ripper-support assembly represents in laboratory environment the behaviour of a bulldozer ripper. A neural network based methodology was tested using experimental data collected from strain gages and a load cell. The obtained results are promising and show a good way to develop a methodology based on neural networks for real-time monitoring of the forces acting on the ripper. This methodology will be extremely useful for the bulldozer operator to know when to stop ripping process.

References

Abstract

In this work we build the first BI-RADS parser for Portuguese free texts, modeled after existing approaches to extract BI-RADS features from English medical records. Our concept finder uses a semantic grammar based on the BI-RADS lexicon and on iterative transferred expert knowledge. We compare the performance of our algorithm to manual annotation by a specialist in mammography. Our results show that our parser’s performance is comparable to the manual method.

1 Introduction

Mammography findings are described by a radiologist according to the Breast Imaging Reporting and Data System (BI-RADS), created by the American College of Radiology [2]. The BI-RADS annotation system specifies a lexicon that forms a common language used by specialists in the area of breast cancer. It is composed of 43 descriptors, organized in a hierarchy. More advanced hospital database systems allow the annotation of BI-RADS features in a standard database format [1]. Nevertheless, most databases either do not support a standard format, or complement it with free-text reports. Hence the need for automated BI-RADS feature extraction from free-text.

Many researchers used natural language processing techniques to extract relevant concepts from medical texts. For example, MedLEE [6] is capable of extracting complex concepts from medical reports written in English. The National Library of Medicine’s Unified Medical Language System (UMLS) compiled a large number of medical dictionaries as well as a vocabulary that specifies a great number of biomedical concepts [8]. Other meta-theauri exist too, like caTIES and SNOMED CT. None of these extracting tools and thesauri incorporate BI-RADS concepts.

In fact, few researchers tackled the problem of BI-RADS feature extraction from mammography free-text reports. Burnside et al. mapped frequent words in medical reports to BI-RADS terms using Linear Least Squares Fit [3]. Even though this approach is language independent, it performs poorly. Nassif et al. used a simple and effective parser, based on regular grammar expressions, to extract BI-RADS terms from English free-text documents [10]. This parser was recently extended to extract breast composition [11].

In this work, we construct a parser to extract Portuguese BI-RADS features, based on [10]. We refine the parser using input from a radiology specialist. We validate our method on a dataset of 153 patients, comparing the algorithm’s performance to manual annotation. Our parser achieves a performance comparable to the manual method performed by a specialist. According to our knowledge, this is the first work on extracting BI-RADS features from medical reports written in Portuguese.

2 Materials and Methods

We used data collected for women undergoing mammography, between 2008 and 2009, at Centro Hospitalar São João in Porto, one of the largest hospitals in Portugal. Data were properly de-identified before the experiments. In total, we had medical evaluations for 622 women and 1,129 mammography reports. After removing redundancies, preprocessing the data, and linking together reports for the same patient, we ended up with 153 instances. Each resulting instance has both kinds of mammography reports: the basic screening and the detailed diagnostic. To best validate our method, we constructed a non-skewed data composed of both types. These 153 dual-instances were also given to a specialist that manually extracted the BI-RADS features after reading each text report.

Our first step was to translate the BI-RADS lexicon to Portuguese. This was done with the help of a specialist. We then built a dictionary of synonyms for every BI-RADS term. Iteratively, we supplemented this list using expert knowledge to differentiate uses of the same word, to gauge medical wording practices and idiosyncrasies [10]. To illustrate some of the terms, the sentence “lesão da pele” (skin lesion) is captured by the presence of both words: lesão (lesion) and pele (skin), within a relative short distance. Thus we established an order for the combination of terms and a degree of proximity. We perform stemming and group words in the same concept if they are synonyms or typos. For example, “adenomegalia”, “axila positiva” and “gânglio axilar” are all associated with the same concept: “Adenopatia Axilar”.

After detecting a concept, we proceed to the treatment of negations. Fortunately, medical texts tend to have a less complex semantic structure, a limited purpose, and are lexically less ambiguous than unrestricted documents [12]. Clinical negations tend to be much more direct and straightforward, especially in radiology reports [9]. A very small set of negation words accounts for the large majority of clinical negations. Following [7], we identify a set of negation triggers: “não” (not) when not preceded by “onde” (where), “sem” (without), and “nem” (nor). We found that negation triggers usually precede, but sometimes fall within or succeed, the concept they negate.
We implemented our algorithm in Perl. The evaluation of the parser was done in 3 phases.

In all phases, we generated a binary matrix of 153 rows by 43 columns, corresponding to the presence (1) or absence (0) of each feature for a given patient. In order to compare the performance of the parser with the performance of the manual annotation, we counted the number of agreements between both, as well as the number of disagreements related to each. We further examined these disagreement cases to determine their correct classification.

3 Results

After going through the 3 phases, combining both data subsets together, our method extracted 206 features, 174 of which are in accordance with manual extraction (84.5%). More details about results of intermediate phases can be found in [4]. Our parser extracted 32 features that the expert didn’t, while the radiologist had 16 extra features. Out of these 48 disputed cases, the parser edges the radiologist by correctly classifying 25 (52.1%). The parser is thus able to discover features missed or missclassified by the radiologist, and exhibits a similar performance.

Figure 1 summarizes the improvements of the parser during the three phases of the experiment, in terms of concordant and discordant extracted features. Each phase is represented by four bars. The first two bars correspond to the screening reports while the next two correspond to the diagnostic reports. Taken in pairs, the left bar (Screening-C and Diagnostic-C) reports the number of concordances between the parser and the radiologist, while the right bar (Screening-D and Diagnostic-D) reports the discordances, features that were either extracted by the parser or by the radiologist but not by both. For the diagnostic reports, we observe a drastic improvement between the first and second phases, and an additional slight improvement by the third phase. For the screening reports, the improvement is not so pronounced, because this type of reports is less thorough and detects less BI-RADS features. In both cases, we managed to achieve a high level of concordance while reducing the number of discordances.

![Figure 1: Number of concordant and discordant extracted features by the parser and the manual methods, over the three phases and both data subsets](image)

The number of features found in the screening mammography reports is much smaller than in its diagnostic counterpart, as expected. However, the parser manages to return approximately the same number of features extracted by the radiologist. The differences in the labeling of features related to shape and margin are explained by sentences containing ambiguous texts about “irregular shape” and “indistinct” or “ill defined” margin. As the expressions “irregular”, “indistinct” and “ill defined” can be used to categorize both shape and margin, the parser ended up extracting more features than needed for shape and margin. In such situations a human inspection of the text can be more effective than the parser. Combining the attributes margin and shape could be a solution, but we would loose granularity, since the BI-RADS lexicon finds it important to distinguish the definitions of margin and shape.

We noticed there is a clear difference in extraction for the concept “Associated Findings”. The parser did not manage to extract all the information available in the medical reports. This difference is due to sentences related to “distorção arquitetural do estroma” (architectural distortion). The different ways of defining the same concept were not well captured by the parser, but were captured by the radiologist. On the other hand, the parser extracted features related to “popcorn” calcifications, while the radiologist missed them.

4 Conclusion and Future Work

Feature extraction from free-text medical reports is still a challenge. In this work, we introduce the first BI-RADS parser for the Portuguese language. We use a simple approach, based on regular expressions, to capture most of the BI-RADS features expressed by radiologists in Portuguese free-text medical reports. We applied our technique to screening and diagnostic mammography reports. Our method is comparable to manual annotation. The parser was in accordance with the manual method over 84.5% of its extracted features, and correctly classified 52.1% of the disputed cases. Our parser may be used as an automated double reader, or, as an assessment tool of radiologist’s labeling of mammography reports. But most importantly, it is an automated method for extracting BI-RADS features from free-text reports and populating structural databases. Breast cancer models and classifiers are built using structured databases [5], and our parser is a necessary step to integrate free-text datasets to such models. Our next step is to apply our parser to other medical reports not used in this study. We are also working on integrating this parser to a medical system that transcribes audio speech into text. Finally, we plan on using the features extracted by the parser to build classifiers that can distinguish between malignant and benign cancer findings in Portuguese medical records.

Acknowledgments

This work has been supported by the projects DigiScope (PTDC/EIA-CCO/100844/2008), HORIZ (PTDC/EIA-EIA/100897/2008), the Fundação para a Ciência e Tecnologia (FCT/Portugal), and the US National Institute of Health grant R01CA127379-01.

References


Abstract

A method for the segmentation of the mammary gland of Breast Ultrasound exams is proposed, resulting in the selection of the Region of Interest for pathological diagnosis. To achieve this, several methods that enhance the transition areas or remove the speckle were selected: Non-linear diffusion, Speckle Reducing Anisotropic Diffusion, Entropy filter, Moran index and Homomorphic filter. Machine learning (SVM and Adaboost) and clustering (K-means) algorithms are used for tissue classification. A segmentation line based on graph theory is computed to finalize the segmentation and a comparison between the classification methods is provided.

The developed segmentation methods were applied to a database with 61 images, 34 without masses and 27 with masses collected using digital support, and segmented by an experienced medical oncologist in Centro Hospitalar da Cova da Beira in Portugal.

1 Introduction

Breast cancer is the leading cause of cancer death worldwide. A gradual decrease of mortality has been registered mainly due to the early detection [1]. Breast Ultrasound (BUS) exams provide a cheap diagnosis tool, without any ionizing radiation, although they are difficult to analyse.

Typical BUS exams are composed of four regions: skin, mammary gland, pectoralis muscle and other tissues (see figure 1). The Region of Interest for pathological diagnosis is the mammary gland. For this reason a reliable segmentation of the gland region from the other BUS image regions becomes important. Most pattern recognition methods applied to BUS exams can benefit of an initial gland segmentation, either if they are used for segmentation, registration or classification [7]. Moreover this separation also might be useful to the radiologist that interprets the exam.

From our knowledge, this problem was also addressed by [3]. Moreover, the work presented in [4] is extended using an adaboost classifier and cross validation.

Figure 1: BUS sample image

2 Proposed method

The proposed method can be separated in 3 steps. In the first step pre-processing algorithms are applied to the BUS image resulting in the motion of speckle and enhancement of the image features, in particular the tissue transitions. Machine learning algorithms or clustering analysis is applied for the tissue classification in the second step. The third step uses the classification result for the segmentation line definition based on a set of heuristic rules and graph theory.

2.1 Image pre-processing

Several methods were applied to the original image in order to enhance the Ultrasound (US) image characteristics. Speckle Reducing Anisotropic Diffusion (SRAD) is a diffusion method designed for ultrasound images [10], that enhances the edges of the image.

The homomorphic filter method normalizes the brightness across an image (low frequencies attenuation) and increases contrast (high frequencies amplification), using a non-linear mapping [5].

Perona and Malik non-linear diffusion method [8] uses a diffusion process based on the heat equation to perform an intra-region smoothing, resulting in an edge enhanced image. The diffusivity constant $\kappa$ is dynamically computed as suggested in [8].

The entropy filter was applied in order to quantify the information on the exam [9]. This feature allows to discriminate the shadows on the image, as the entropy value is very low on these regions.

Finally, the Moran index measures the unsharpness of a region (window), resulting in an image with enhanced contours. The smooth regions are represented with low values and the heterogeneous zones are represented with higher values like the non linear diffusion.

The bins of the pixel descriptor used on the classification step is defined by the result of the pre-processing algorithms.

2.2 Tissue classification

The tissue classification was done using three different approaches. A first approach was done using the K-means clustering algorithm. A 4 bin pixel descriptor was used because the entropy bin was not included. The output of this method is an image with K clusters (values of $K=3$, 4, 5, 6 and 7 were used), represented with different gray levels.

The second and third approach use respectively the SVM and Adaboost machine learning algorithms, using the 5 bins descriptor defined before. Both methods require a supervised learning, being the training set randomly selected and using the remaining images as test set. A binary image results of these tissue classification methods.

2.2.1 Binary Mask Images

Using the defined image pixel descriptors, binary images are computed resulting in a rough tissue separation. The K-means results in an image with K different levels. Selecting each K level as foreground, and all the others as background, results in K binary images.

The binary images that result from the classification algorithms are then post processed by the algorithm described in subsection 2.3 and 2.4 to result in a final segmentation line.

2.3 Segmentation line definition

Graph theory [2] is used to define the segmentation line defining the shortest line between the left and right side of the binary image. Initially, the horizontal Sobel operator is applied to the image, resulting in a binary image with a set of edges. This step suppresses the vertical edges, keeping only edges with horizontal component.
Table 1: Segmentation results obtained using K-means pixel classification in images without masses.

<table>
<thead>
<tr>
<th>K</th>
<th>(l_{th})</th>
<th>TP(%)</th>
<th>TN(%)</th>
<th>FP(%)</th>
<th>FN(%)</th>
<th>Acc. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>95.53</td>
<td>70.22</td>
<td>29.78</td>
<td>4.47</td>
<td>82.87</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>95.03</td>
<td>51.94</td>
<td>48.06</td>
<td>4.97</td>
<td>70.31</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>95.44</td>
<td>76.41</td>
<td>23.59</td>
<td>4.56</td>
<td>71.14</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>95.20</td>
<td>72.47</td>
<td>27.53</td>
<td>4.80</td>
<td>78.91</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>98.83</td>
<td>64.76</td>
<td>35.24</td>
<td>3.17</td>
<td>79.72</td>
</tr>
</tbody>
</table>

Table 2: Segmentation results obtained in BUS exams using SVM.

<table>
<thead>
<tr>
<th>Without masses</th>
<th>TP(%)</th>
<th>TN(%)</th>
<th>ACC(%)</th>
<th>(\sigma_{AC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>92.51</td>
<td>93.10</td>
<td>91.81</td>
<td>7.55</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With masses</th>
<th>TP(%)</th>
<th>TN(%)</th>
<th>ACC(%)</th>
<th>(\sigma_{AC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.60</td>
<td>86.61</td>
<td>84.80</td>
<td>10.60</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Segmentation results obtained in BUS exams using Adaboost.

<table>
<thead>
<tr>
<th>All database</th>
<th>TP(%)</th>
<th>TN(%)</th>
<th>ACC(%)</th>
<th>(\sigma_{AC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.12</td>
<td>90.42</td>
<td>88.00</td>
<td>10.60</td>
<td></td>
</tr>
</tbody>
</table>

A label is then applied to each different horizontal edge. A weighted undirected graph is created using as nodes the edges (labels) and as weights the minimum Euclidean distance between each edge (label). Using the resulting graph, the shortest path between the left and right side of the exam is computed. The selected edges that result in the shortest path are linked with a line connecting the extremes. The final segmentation line is defined by the three shortest paths. The top of the exam (rows) was ignored on the edges labelling, avoiding the definition of the segmentation line on transition between the skin/gland that has similarities with the gland/muscle transition.

### 2.4 Segmentation line selection

The K-means clustering algorithm results in K binary images (one for each cluster). After applying the separation line algorithm previously described, 3 × K possible separating lines result. For each separating line, a binary mask is created, with value 1 above the separating line and 0 below. Subsequently an element-wise sum is applied to the generated masks, resulting in an image with (3 × K + 1) levels. The separation line is selected applying a threshold, \(l_{th}\), to the element-wise sum image. The threshold \(l_{th}\) is selected based on the value that provides the overall best results for our database (Accuracy Vs. False negatives) and is a function of the K value.

For the binary image that results from the support vector machine (SVM) and Adaboost classification, 3 possible separation lines are returned. If some of the lines overlap, the shortest one (in length) is selected. If a separation line with no overlap occurs, the bottom most line is selected, either it is an overlapping one or not. The option for the bottom most is related with the fact that this separation line is likely to be the one that guarantees a higher probability that all pixels of the gland regions are inside the top region. Hence, a reduced amount of pixels of the BUS images gland is suppressed by this segmentation method, resulting in a higher Recall rate with a cost of lower Precision rate.

### 3 Testing and results

#### 3.1 Database description

The used database was collected at “Hospital da Cova da Beira”, in Portugal, using a LOGIQ Book XP. The database is composed by 61 images, 34 without masses and 27 with masses, collected in digital support without losses (lossless compression). The images without masses have a typical size of [412 × 391] pixels. The images with masses the number of columns is typically the same (412) and the number of rows varies from 391 to 524. An experienced medical oncologist on US image analysis did the tissue separation for validation, resulting in masks of the gland tissue created manually.

#### 3.2 Parameters definition

Relatively to the image pre-processing and after intensive testing several parameters were settled:

1. The entropy and Laplacian filters pre-processing algorithms use a seven pixels square window.
2. In the Perona and Malik non-linear diffusion [8] the number of iterations was set to 25, the \(\kappa\) to 60% and the \(\lambda\) was set to 0.25. Moreover, the Gaussian diffusity function was selected.
3. In the SRAD algorithm [10] the number of iterations was set to 20 and the smoothing time step (\(\lambda\)) was set to 0.05.
4. The homomorphic filtering [6] used a boost value of 2, a CutOff of 0.4, a \(\text{histogram\_cut}\) of 3.75 and a \(\text{histogram\_cut}\) of 5 with a filter order of 3.

The segmentation line definition for K-means, 70 rows were ignored on the top of the image.

### 3.3 Results discussion

The presented results show that the proposed task can be achieved with a high success rate. When analysing the results from the K-means approach the accuracy rate, on the best results (K = 3, 6 and 7), is between 78.91 and 82.87. This approach results in a TP rate over 95%, that however has an unacceptably high FP rate. The machine learning approach revealed to be much more accurate. Besides the lower TP rate, that we consider acceptable, the accuracy rate (86.91 and 91.81) and TN rate (87.48 and 93.10) are much higher.

The results on table 2 reveal better results on the segmentation on images with masses than on images without masses. However no cross validation was done, and these results are very dependent of the randomly selected training sets.

The Adaboost results were obtained using cross-validation with random training sets selection. These segmentation results are possible to be used as a pre-processing method for any further BUS exam analysis, since it accurately remove the unnecessary parts of the exam.

### References


Extracting Health Information From Portuguese Tweets

José Carlos Santos  
josecarlos@ua.pt

Sérgio Matos  
aleixomatoss@ua.pt

IEETA  
University of Aveiro  
Aveiro, Portugal  
http://bioinformatics.ua.pt

Abstract

Social media platforms encourage people to share diverse aspects of their daily life. Among these, shared health related information might be used to infer health status and incidence rates for specific conditions or symptoms. In this work, we evaluate the use of Twitter messages and search engine query logs to estimate the incidence rate of influenza like illness in Portugal.

Based on a classified dataset of 2704 tweets from Portugal, we obtained a precision of 0.86 and an f-measure of 0.75 using an SVM classifier. Using feature selection based on mutual information, we achieved a precision of 0.80 and an f-measure of 0.84 for a Naïve Bayes classifier with 650 textual features. Combining this classifier with the frequency that influenza related words are searched on the Internet, we create a timeline that estimates the incidence of flu in Portugal. Comparing the results to epidemiological data over a time period not considered during training, a correlation coefficient of 0.84 was obtained. Although the Portuguese community in Twitter is small, our results are comparable to previous approaches in other languages, and indicate that this approach could be used in the future to complement other measures of disease incidence rates.

1 Introduction

Social media platforms such as Twitter encourage people to share their opinions, thoughts and life aspects [8]. Among these, people often share personal health related information, such as the appearance of flu symptoms or the recovery of those symptoms. Thus, this information could be used to identify flu cases and estimate the influenza rate through time.

Several works regarding the retrieval of health information from social media have already been published. E. Aramaki et al. [1] applied SVM machine learning techniques to Twitter messages to predict influenza rates in Japan. Lamps and Cristianini [6] and A. Culotta [2, 3] analysed Twitter messages using regression models, in the United Kingdom and the United States respectively, obtaining correlation rates of approximately 0.95. Different works also rely on query logs to track influenza activity. Ginsberg et al. [5] presented Google Flu Trends, which uses Google search queries and achieves an average correlation of 0.97. The greatest advantage of these methods over traditional ones is instant feedback. Health reports are published in a weekly or monthly basis, while Twitter, and/or query log analyses can be obtained almost instantly. This characteristic is of extreme importance because early stage detection can reduce the impact of epidemic breakouts [1, 5]. In this work, we integrate information from tweets and from query logs in order to estimate the occurrence rate of influenza in Portugal.

2 Methods

We obtained training data from nearly 14 million tweets originated in Portugal and covering the period between March 2011 and February 2012. We excluded re-tweets (replies) and tweets including links. Using a regular expression that captures tweets that have influenza related words, a set of 3183 tweets was identified. The regular expression used was divided into three groups: “gripe” (flu) word derivations, “constipação” (cold) word derivations and flu related symptoms, such as body/throat pains, headache and fever, as described in Table 1.

Using filtering based on regular expressions of this kind is not sufficient, as many tweets that contain words related to flu do not imply that the person writing the text has the flu. Tweets like “Hoping the flu doesn’t strike me again this Winter” contain the keyword flu but do not tell us that this person has the flu. To solve this problem, machine learning techniques were used. We manually classified a set of tweets as positive or negative to be able to train a predictive model. Tweets were labelled as positive if they revealed that the person who wrote it was with the flu, was having flu symptoms or was recently ill with the flu. A total of 2704 tweets were classified, of which 949 were positive for flu.

2.1 Machine Learning Methods

Tweets were represented by a bag-of-words (BOW) model. The Natural Language Processing Toolkit [9] (NLTK) was used to tokenize the text, remove portuguese stopwords and stem all remaining words in each tweet. Char bigrams for each word were also generated, making up a total of 5106 features.

Several machine learning techniques (SVM, Naïve Bayes, Random Forest, Classification Tree, Nearest Neighbour) were tested in order to evaluate which would produce better results. We used the SVM-light [7] implementation of SVMs. The remaining classifiers were trained using the Orange toolkit [4].

2.2 Feature Selection

In order to improve these results, we evaluated the use of feature selection techniques for defining the best set of features to use. For this, each feature was compared to the true class label set to obtain the mutual information (MI) value. The higher a feature’s MI score, the more related it is to the true class label, meaning that the feature contains discriminative information to decide if that tweet has a positive or negative annotation. We selected the optimal number of features empirically, by varying the threshold on the MI value and selecting features with MI above that threshold.

2.3 Query Logs

Besides Twitter, we also had access to query logs from the SAPO1 search platform. We used around 15 million log entries from December 2011 to May 2012, which gave us an average of 112 thousand log entries per day. We filtered all searches using the regular expression (en)?grip[a-z]+, which matches flu influenza word derivations. The number of search queries matching this regular expression was accumulated weekly.

3 Results

3.1 Binary Classification of Twitter Messages

The performance of the different classifiers was compared through 5×2-fold cross validation using the entire dataset of 2704 tweets. Using the full set of features (Table 2), the best results were obtained with the SVM classifier, with an f-measure of 0.75. After feature selection (Table 3), the best overall results were obtained for a set of 650 features, achieving an f-measure of 0.84 with the Naïve Bayes classifier.

For each classifier, we selected from the receiver operating characteristic (ROC) analysis, an operating point that maximizes the precision without a severe loss on the classifier recall. Although operating points

<table>
<thead>
<tr>
<th>Theme</th>
<th>Regex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flu</td>
<td>(en)/grip[a-z]+</td>
</tr>
<tr>
<td>Cold</td>
<td>constip[a-z]+</td>
</tr>
<tr>
<td>Flu Symptoms</td>
<td>(febre .* grau(s)?)</td>
</tr>
<tr>
<td></td>
<td>(bodypains, * jebre)</td>
</tr>
</tbody>
</table>

Table 1: Regular expressions to filter influenza related tweets.
with higher f-measure values could have been selected, these would represent higher recall, at the expense of a lower precision. We therefore chose the more stringent models, in order to reduce the amount of false positive hits, and consequently, the amount of noise present in the final results.

### 3.2 Flu Trend Prediction

As stated before, the main objective of this work is to create a reliable predictive model to obtain instant feedback regarding the incidence of flu in Portugal. In order to access the performance of our approach, we compare it to epidemiological results from a health-monitoring project related to flu. The gold standard data is collected from the Influenzanet\(^2\) online system. Influenzanet data is collected from several participants who sign up to the project and report any influenza symptoms, such as fevers or headaches on a weekly basis. The gold standard period considered runs from 28 November 2011 to 22 April 2012. Pearson’s correlation coefficient is used as the evaluation metric between the gold standard and the predictor output.

In order to compare the tweet and query results to the flu incidence rates, these were expressed as weekly relative frequencies, as done in previous works. This value is calculated as the fraction of tweets (searches) considered positive to the total of tweets (searches) produced during each week. The data were time aligned with the Influenzanet results.

We compared the generated output of our classifiers with the gold standard timeline. As shown in Table 3, we obtain the best results when we combine the Naïve Bayes classifier results with the result from the query logs analysis. Since both are represented in terms of relative frequency occurrence, we simply add those values. The Naïve Bayes classifier produced the higher correlation coefficient, in accordance with the binary classification performance.

### 4 Discussion

To the best of our knowledge this is the first work in this subject done specifically for the Portuguese language. Although most of the used methods are similar and applicable across languages, the amount of available data in languages other than English, as well as language specificities, may influence the final results obtained.

Despite Twitter being a largely used social web platform, it is not very popular in Portugal, which limited the size of our dataset when compared to similar works. As a comparison, we had access to around 14 million tweets, with a daily average of nearly 40,000 tweets, from which 2704 were used to train the binary classifiers. Eiji Aramaki et al. [1] used 300 million tweets, from which 5,000 were used for training. A. Cullotta [2] used a total of 500,000 messages, selecting 206 of those messages to train a model. Due to the low set of used data, overfitting problems are reported in that work.

Another important novelty of our work is the integration of results from the analysis of tweets with results from user searches on a web search engine. This contributed to a better approximation to health monitoring results used as gold-standard in this work. A possible extension to this would be to user other sources of user-generated content, as blog posts and comments on web pages.

The results obtained indicate that this method can be used to complement other measures of disease incidence rates.

![Figure 1: Influenza rate from December 2011 to April 2012 showing the results from the Influenzanet project and the combined result of our Naïve Bayes classifier with the query log rate. The correlation ratio between the two sets of measures is 0.842. Naïve Bayes classifier was trained with data from the period 01/05/2011 to 03/12/2011.](image)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>F-Measure</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>0.66</td>
<td>0.81</td>
<td>0.55</td>
</tr>
<tr>
<td>SVM</td>
<td>0.75</td>
<td>0.86</td>
<td>0.66</td>
</tr>
<tr>
<td>Classification Tree</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.68</td>
<td>0.70</td>
<td>0.67</td>
</tr>
<tr>
<td>kNN</td>
<td>0.78</td>
<td>0.67</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classifier</th>
<th>F-Measure</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>0.84</td>
<td>0.80</td>
<td>0.90</td>
</tr>
<tr>
<td>SVM</td>
<td>0.78</td>
<td>0.84</td>
<td>0.72</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.79</td>
<td>0.72</td>
<td>0.86</td>
</tr>
<tr>
<td>Classification Tree</td>
<td>0.76</td>
<td>0.74</td>
<td>0.77</td>
</tr>
<tr>
<td>kNN</td>
<td>0.79</td>
<td>0.71</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 3: Classifier results after feature selection (650 Features)

|| Dataset          | SVM (p-value) | Naïve Bayes (p-value) |
|-------------------|---------------|-----------------------|
| Twitter           | 0.61 (25.75 \times 10^{-5}) | 0.68 (4.98 \times 10^{-4}) |
| Twitter + Query Logs | 0.75 (0.58 \times 10^{-4}) | **0.84 (0.01 \times 10^{-4})** |

Table 4: Pearson’s Correlation Ratios to incidence rates from Influenzanet

### Acknowledgements

This work was funded by Labs SAPO project “SPoTED”.

### References


\(^2\)http://www.influenzanet.eu
Simulation of Ultrasound Guided Robotic Surgery with MORSE

Ruben António
rubenmcantonio@gmail.com

P.J.S. Gonçalves and P.M.B. Torres
paulo.goncalves@ipcb.pt / pedrotorres@ipcb.pt

Polytechnic Institute of Castelo Branco, School of Technology,
Av. Empresário, 6000-767 Castelo Branco, Portugal

Abstract
The first decade of the 21st century has seen a widespread deployment of Robotic Surgery Systems for use in a growing range of complex surgical procedures. To perform tasks in surgeries, namely navigation and object recognition, vision systems are of great importance to increase autonomy of surgical systems. Ultrasound (US) images are of special interest in the medical field because they are radiation free and cheaper than computer tomography (CT), or magnetic resonance (MR), despite its intrinsic speckle noise. Advanced materials and technology have a significant cost. With this scenario, simulators gain even more importance, providing essential knowledge about the systems behaviour before being placed in the real working scenarios. Our purpose is to use the Modular OpenRobots Simulation Engine (MORSE), a robotics simulator, to create a simulated surgery room environment and a constant flow of data between the simulator and other equipment and software’s that allow us to test the performance of Ultrasound Guided Robotic Surgery Systems.

1 Introduction
Robotic surgery was developed to enhance the capabilities of surgeons performing open surgery and to overcome the limitations of minimally invasive surgery. This technology allow the surgeons to perform certain actions with smoother, feedback-controlled motions than could ever be achieved by a human hand, reduce or eliminate the tissue trauma associated with open surgery and the ability to perform a remote surgery where the physical distance between the surgeon and the patient is immaterial. Robotic surgery systems are becoming highly complex and sophisticated, with an increasing number of hardware and software components. The common use of images to control robotic systems, i.e., placed directly in the control loop, also being deployed to the medical field. Applications started using standard images in open surgery, but the need to “see” inside the human body lead to the use of medical images, e.g., US, MR or CT. This paper is tackled the case of ultrasound images for the visual feedback of a robotic manipulator performing a orthopaedic surgical task. There are also a large variety of tasks involved in performing robotics experiments, which needs much time and resources for validation. The use of a simulator can ease the development and validation processes, allowing to verify the component integration and to evaluate their behaviour under different controlled circumstances. Simulation is cheaper in terms of time and human resources than experiments with real robots and medical imaging equipment. Another advantage is that a simulated environment can be significantly more complex and larger than a lab environment, and meanwhile ensure a perfect repeatability. Nowadays there are several robotic simulators, like for example Gazebo, MORSE [1], Webots or Microsoft Robotics Developer Studio. After analysing those software’s we concluded that MORSE was our best choice because it is an open-source application that can be used in different contexts for testing and verification of robotics systems as a whole, at a medium to high level of abstraction. One of the main interests of the simulator is that it can be reusable by researchers, engineers and is being upgraded as part of multiple projects. Another important thing is that the simulator can interact with any middleware used in robotics, e.g., ROS [2] or OROCOS [3], and not just impose a format that others must adapt to.

2 MORSE robotic simulator
The MORSE simulator is an open-source application, that is built on top of Blender, an open-source 3D modelling and rendering application, using its powerful features and extending its functionality through Python scripts and it can be used in different conditions and scenarios for the testing and verification of robotics systems. The simulator contains a library of simple components that can be assembled together with others, and each component consists of a Python and a Blender file. The Python file defines an object class for the component type, with its state variables, data and logical behaviour or methods. The Blender file specifies the visual and physical properties of the object in the simulated world. There are currently three different kinds of robotics components defined in MORSE. The first are the sensors that can recover data from the simulated world and emulating the functionality of the real sensors. Then we have the actuators that produce actions upon the associated robots or components. Finally are the robots where the sensors and actuators are mounted. Besides those, there are other classes we can use like scenes, middleware and modifiers. The scenes are modelled environments, created with Blender, where the robot will interact during the simulation. Middleware is an intermediate layer between the simulator and different software systems that enables them to communicate and share data. In order to approach the values obtained in the real world we can use modifiers that have the role to alter the data produced by the simulator, mostly by using noise functions. Our objective is to use these tools in order to simulate a surgery room environment with an ultrasound guided robotic system designed for orthopaedic surgery. Other important step is to recreate in Blender a 3D bone registration module, between intraoperative 3D point data from US images, and pre-operative 3D point data obtained from CT images. The 3D registration module that we are going to use its obtained through bone contour segmentation from ultrasound images and afterwards, using point cloud algorithms the 3D registration is performed.

To keep in track every component position or motion and receive or send data in real time we require additional software linked to MORSE. To achieve this we are going to use a specific middleware, OROCOS connected with Robot Operating System (ROS), both open-source software frameworks for robot software development, that will manage the intermediate layer between MORSE, Blender and Linux. In Figure 1 is depicted how ROS manage the data flow in the intermediate layer, between MORSE and Blender, during the simulation.

![Figure 1: Simple demonstration of how ROS manage the data flow in the intermediate layer, between MORSE and Blender, during the simulation.](image-url)

For the ROS management, we are going to use the rospublish-subscriber model, where each node is interested in a certain kind of data will subscribe to the appropriate topic, which is simply a string such as “odometry” or “map.” A node that is interested in a certain kind of data will subscribe to the appropriate topic. The topic-based publish-subscribe model is called service. The
service is defined by a string name and a pair of strictly typed messages, one for the request and the other for the response.

3 Ultrasound Guided Robotic System

In this section is depicted the software tool (SWT) that is used for the MORSE simulation and afterwards will be used for the real-time control of the system.

![Software Tool Architecture diagram](image)

The Software Tool architecture is depicted in figure 2. The libraries IGSTK [4], OpenCV [5], OROCOS/ROS and PCL, play a decisive role in the interaction between the SWT and the physical devices: US probe, NDI Polaris Tracker, Robot Manipulator and the visualization hardware. The SWT is multi-platform, although we are using UBUNTU, linux. IGSTK (Image-Guided Surgery Toolkit), sets the bi-directional communication with the Polaris Spectra and with OpenCV. The later is responsible for acquiring and processing the acquired US images. These two libraries are responsible for the calculation of the 3D Point Clouds, from the US probe and Polaris data. Through the function igstk::PolarisTracker inserted in the class igstkNDITracker, it is possible to communicate with the Polaris Spectra system, start communications and store data of the location of the probe in the 3D space. OpenCV is responsible for the acquisition of the US images, that arrive by a off-the-shelf video card, placed in the workstation. Using classes and features from OpenCV the 3D Point Cloud is obtained using algorithms based in image segmentation and tracking of the contour bone, that are implemented in the SWT. It processes each US image, locates them in 3D space with information that arrives through IGSTK and follow the movement of the bone, based on the information of its contour in each US image. The SWT creates a 3D point cloud, to be processed by PCL, that is responsible for visualization and the registration between the intra-operative and the bone model 3D point clouds.

The OROCOS project is an Open Robot Control Software that have ROS bindings to interact with MORSE and the simulation scenario therein, developed in C++, with a very important role in the SWT, i.e., it can be easily simulated and deployed to the robot controller. It is responsible for controlling the robot, based on the information gathered from the surgical environment. If the position and translation of the femur changes during the surgery, the robot manipulator will compensate for the pre-operative position and orientation of the femoral head, and be correctly positioned to perform a drilling task.

4 The Simulated Scenario

This section presents the first results obtained at the current developing stage of the simulations. At this point and regarding figure 2, the SWT have a stable link between all modules. The robot manipulator used for the simulations is a KUKA LWR arm actuator, the orange robot manipulator in figure 3. It is a robotic arm composed of 8 segments and an armature that controls its movement. This actuator reads a set of seven joint angles to move the robot links, obtained from the robot inverse kinematics of the drilling point of the femoral head referred in the last section, and obtained from the US images and the Polaris tracker. The armature can also send data about the minimum and maximum angles that any segment can rotate to, i.e., to define the robot workspace. With this data we are going to program the arm to iteratively adjust the segments to prevent that, in any joint rotation command, none of the segments reaches its maximum rotation before completing the sequence, or auto collide. Its essential to use position sensors to determine if the position of the segments moving parts, is in the robot workspace. For example we must know at all time where is the end-effector of the KUKA LWR, and the tool attached to its end-effector.

![MORSE scenario with the KUKA LWR arm appended to the Ido robot performing segment rotation.](image)

After creating the surgery room environment, depicted in figure 4, and the Ultrasound Guided Robotic Surgery System is installed, our goal is to create a Python script, which will run on Python3 through the Linux console, with all commands available for the robotic arm and sensors, i.e., Polaris and the US probe.

![Development of the surgery room environment.](image)

5 Conclusions

Nowadays, simulating robotic systems is essential, specially the ones with medical applications due to its complexity and difficulty to work with humans or cadavers. MORSE is a useful simulator for this task. It is open-source software, which is rapidly evolving and can change its behaviour according to the user needs. When controlling the KUKA LWR it was verified that the simulator has a huge potential and it's very precise executing tasks, presenting smooth trajectories when using top-end graphic cards. When using ROS as the middleware, the communication between several modules, working in a second computer, is performed using sockets. The ultrasound probe and the Polaris tracker are connected to the second computer, that processes the signal and sends real data to the simulator, responsible to perform the simulated robotic surgery. The deployment of the simulation to the real robot controller can be performed using OROCOS or the FRI, i.e., the KUKA control interface.

References

Abstract

This paper addresses the problem of providing effective situation awareness to the operator of a remote vehicle. To this end, we propose the use of a stereo camera pair, mounted on a pan and tilt servo system, together with a Head-Mounted Display (HMD) worn by the operator. The HMD is equipped with a head tracker, providing the attitude of the operator head. The goal of this system is to stream into the HMD images taken by the stereo cameras, such that this stream is perceived with the same attitude (apart from a fixed rotation along yaw axis) than the one of the operator head. This is accomplished by an hybrid approach, combining pan and tilt movement of the servo system, with a virtual pan, tilt, and roll by rectification of the images. This rectification, on the one hand, overcomes the dynamical limitations of the servos, in terms of responsiveness, and on the other, introduces roll rotation of the images (not possible since the camera mounting does not possess that degree of freedom). This arrangement allows not only an immersive experience to the operator, but also decouples the kinematic control of the vehicle from the camera attitude control. A prototype of the complete system is currently in operation.

1 Introduction

The utilization of search and rescue robots in danger scenarios is becoming a more and more common practice with the advance that this area have faced through the last decade. These robots are often remotely operated by humans. The operator task is quite difficult: besides paying attention to the sensors information while driving the robot, the operator must fully understand the environment around the robot and the consequences of his decisions concerning the robot motion. In order to enhance the operator perception of the environment, we propose to use a 3D vision system.

The system is comprised by an Head Mounted Display (HMD) that provides 3D vision to the robot operator. The 3D images are captured by a stereo camera mounted on a pan and tilt system (figure 1). The HMD contains an head tracker, providing the operator head attitude (figure 2). With this setup the pan and tilt system will follow the head attitude, providing a more intuitive control of the camera orientation. Studies have shown that this setup enhance the operator situational awareness, granting a more safe control of the robot [3]. The main problem of this setup is the pan and tilt motor responsiveness. This response tends to be slow for large changes in the camera orientation, which results in a lagged experience for the operator. Also, generic pan and tilt systems have low resolution stepper motors, that may lead the operator to experience some flicker in the image for small pan and tilt movements. To overcome these limitations a virtual pan, tilt and roll method is applied to the stereo images. This method combined through a controller, with the mechanical servos position, provides a more fluid experience to the operator. This work was developed in RAPOSANG-NG1, a tracked wheel robot designed for search and rescue operations.

2 Related Work

Virtual pan, tilt and zoom to simulate real pan, tilt and zoom cameras is a well discussed subject [4]. It is achieved by selecting a Region Of Interest - ROI - from a wide and high resolution image. Moving this ROI through the image according to a user input provides the user with a pan and tilt feel. If the ROI size is adjustable then the user experience a zoom in/out effect. The main advantage of virtual pan and tilt is that the camera keeps still while the user navigates through the scenario. Although this approach provides a fluid experience to the operator, it fails in providing the right perspective if the scenario is not far enough. The implementation of stereo vision through an HMD for search and rescue robots immersive teleoperation, have shown big improvements in the operator situational awareness [1, 3].

3 Controller Design

The proposed controller is presented in figure 3. It can be divided in two major blocks: the (1) Virtual Pan, Tilt and Roll block, VPTR, where the virtual pan, tilt and roll method is applied to the stereo images, and (2) the mechanical pan and tilt system controller.

Let \( \theta_{HMD} \) be the head mounted display attitude and \( \theta_{PPT} \) the physical pan and tilt orientation, both in roll-pitch-yaw angle representation \((rpy)\). The VPTR block compensates the error between both orientations, \( \Delta \theta \), through the method presented in the next section. The error \( \Delta \theta \), in \( rpy \), can be defined has

\[
\Delta \theta = \theta_{HMD} - \theta_{PPT} = [\Delta \theta_r, \Delta \theta_p, \Delta \theta_y]^T
\]

(1)

The pan and tilt controller comprises a cascade of blocks: an hysteresis, a proportional block, an integrator and a saturation block. The hysteresis
block applies a non-linear transformation in the error $\Delta \Theta$ providing two work modes: the active mode and the idle mode (figure 4). In the active mode the hysteresis block output, $H(\Delta \Theta)$, is equal to the input $\Delta \Theta$. In the idle mode the output of the controller is zero and only the VPTR block compensates $\Delta \Theta$. The switch between modes happen when:

- $-i_a < \Delta \Theta a < i_a$ the controller enters idle mode,
- $-a_a > \Delta \Theta a > a_a$ the controller enters active mode,

with $\alpha \in \{r, p, y\}$. Also, $a_a$ and $i_a$ are adjustable, to provide a smooth transition between hysteresis modes in all the orientations. The pan and tilt system does not provide roll movement, so the roll error is only for VPTR. The hysteresis idle mode prevent small head movements induce a motor movement, being the image perspective corrected by the VPTR. The active mode enables the controller to change the camera orientation while $|\Delta \Theta| > i$. The proportional block converts the error $\Delta \Theta$ into motors velocity $v_{ref}$. The saturation block imposes a limit to the velocity reference of the servos. The integrator converts $v_{ref}$ into a position reference, since the servos are controlled in position. These blocks sequence make possible to use the position sent to the motors as an estimate has $\Delta \theta_{ref}$ because the saturation block prevents velocities higher than the motors maximum speed, granting that the motor position sent is always achieved.

where the $vc$ index stands for virtual camera and $rc$ index for real camera. The relation between $\tilde{w}_{vc}$ and $\tilde{w}_{rc}$ is then given by

$$\tilde{w}_{rc} = \lambda Q_{rc} Q_{rc}^{-1} \tilde{w}_{vc}, \quad \lambda = \lambda_{vc}/\lambda_{rc} \quad (4)$$

The virtual and the real cameras are the same, meaning that both have the same intrinsic parameters, thus $Q_{rc} = A_{rc} R_{rc}$, where $R_{rc}$ is the rotation matrix between the real camera and the virtual camera. If the real camera and the world share the same reference then $R_{rc} = I$, the identity matrix. Then equation (4) becomes

$$\tilde{m}_{rc} = \lambda A R_{rc} A^{-1} \tilde{m}_{vc} \quad (5)$$

Given a rotation matrix between virtual and real cameras, equation (5) relates the real image pixel coordinates with the virtual image pixel coordinates. As this method projects the image through a new orientation provided by the rotation matrix, the image perspective changes, and the user perceives like he is moving the real camera orientation. For the controller presented in the previous section, the rotation matrix is calculated from the error $\Delta \Theta$, providing to the operator the correct perspective for his head attitude. As one can see in figure 6 although the image used to apply the virtual pan, tilt and roll method was taken with a different camera angle, the angles between the objects in both images are very similar.

![Figure 4: Hysteresis block transfer function. In red (where $H(\Delta \Theta) = \Delta \Theta$) is represented the Active mode and in green (between $-a$ and $a$, where $H(\Delta \Theta) = 0$) the Idle mode.
](image)

![Figure 5: Virtual pan, tilt and roll geometry. The virtual camera image, in dashed line, is a rotation of the real camera image over the optical optical center C.
](image)

![Figure 6: Top Left image: taken with real camera tilt orientation of $22^\circ$. Top Right image: taken with the real camera tilt orientation of $0^\circ$ and then applied the virtual pan, tilt and roll method with a tilt angle of $22^\circ$. Bottom image: In red are the pixels origin position and in blue their destination.
](image)

![Figure 5: Virtual pan, tilt and roll geometry. The virtual camera image, in dashed line, is a rotation of the real camera image over the optical optical center C.
](image)

## 4 Virtual pan, tilt and roll

The virtual pan and tilt is achieved by projecting the stereo images through a new orientation. Applying the pin-hole camera model (figure 5), the line that connects a point in the world, $W$, to the camera optical center, $C$, can be parameterized [2] apart a constant $\lambda$, in homogeneous coordinate by

$$W = C + \lambda Q_{rc}^{-1} \tilde{w}_{rc} \quad (2)$$

where $\tilde{w}_{rc}$ is the coordinates of the image pixel in homogeneous coordinates. $Q_{rc} = A_{rc} R_{rc}$, is the product between the intrinsic parameters matrix $A_{rc}$ and $R_{rc}$, the rotation matrix, between the world frame and the real camera frame. The virtual camera orientation is achieved by a rotation of the real camera orientation over $C$. As both virtual and real cameras share the same optical center $C$, to the same point $W$ one can write for both virtual and real cameras

$$\begin{cases} W = C + \lambda_{vc} Q_{rc}^{-1} \tilde{w}_{vc} \\ W = C + \lambda_{vc} Q_{rc}^{-1} \tilde{w}_{rc} \end{cases} \quad (3)$$

## 5 Conclusions and Future Work

In this paper we propose a controller to integrate virtual and real camera controls in order to improve the situational awareness of a search and rescue robot operator. The controller have been implemented in RAPOSA-NG. It showed to be comfortable and to provide an immersive perception of the environment while teleoperating the robot. The virtual pan, tilt and roll method achieves it’s purpose providing the rectified image perspective to the operator. In recent future, a case study will be done in order to confirm these results.

### References


Abstract

Hepatic steatosis is a common histological feature in chronic hepatitis C (CHC). In this paper, some experiments are made to test the usefulness of monogenic filter based features in the discrimination of hepatic steatosis from ultrasound images and to study the use of the Bayes factor as a tool to evidence the presence of hepatic steatosis in patients with CHC. The effectiveness of our method shows the usefulness of the extracted features, achieving an overall accuracy of 93.33%, in the detection of steatosis from ultrasound images and the potential for physicians to quantify the evidence of liver steatosis in CHC patients.

1 Introduction

Fatty liver infiltration, called hepatic steatosis (HS), occurs when the fat content of the hepatocytes increases [13]. Worldwide, the prevalence of HS has increased, in great part associated with obesity, insulin resistance and alcohol [2].

HS is a common histological feature of chronic hepatitis C (CHC) [1] and it has been shown that CH can directly induce HS. CHC is a more severe chronic liver disease with high mortality and morbidity rate. In patients with CHC, necropsy studies detected HS in 31% to 72% of cases, and liver biopsies have demonstrated the presence of HS in approximately 50% of the cases [9]. Compared with the general population, patients with CHC have 2.5 times higher prevalence of HS [14].

Liver biopsy remains the reference exam for HS [2]. The need for biopsy reduction has led to the development of non-invasive methods. Among the new techniques for HS assessment, imaging methods are widely reported in the literature. Ultrasound (US) is a first line tool due to its non-invasive, low price, accessible, and non-ionizing nature.

HS leads to an increase of brightness and textural characteristics changes in the US image. However, the source of these changes may be in reflective echoes arising from other hepatic diseases, such as cirrhosis [12]. Care should be taken in a direct classification decision between HS and CHC/cirrhosis classes. The difficulty in a classification decision between HS and CHC/cirrhosis classes is expressed in several studies [7, 8, 15]. Due to the high probability of CHC patients develop HS these classes should not be treated as independent cases.

A large spectrum of studies has debated the optimal and sub-optimal features extracted from US images for HS classification purposes. In this study, we introduce the monogenic filter as a feature extraction algorithm.

The novelties of this study are the investigation of the presence of HS in CHC patients, based on the evidence provided by the Bayes factor; and the extraction of monogenic filter features for the classification of HS.

1.1 Problem Formulation

In this study, the problem is formulated in two phases: I) Establishment of baseline values extracted from the classification normal versus HS classes; and II) testing the algorithm in patients with CHC and compare the obtained Bayes factor values with the baseline.

In the first phase, we find the feature set that best discriminates HS from normal livers, train it with the Bayes classifier and calculate the Bayes factor to create the baseline values.

A novelty of this study is the extraction of features from the implementation of Felsberg’s monogenic filters (MF) [3] based on the extension to 2D proposed by [6]. Three filter scales are used in this study. From each filter level scale we obtain: a) monogenic Amplitude response (A);

b) phase orientation response (θ); and c) phase angle response (ψ). Given a two dimensional signal \( f(x) \), \( x \in \mathbb{R}^2 \), Felsberd and Sommer [3] define the three-component monogenic signal as

\[
\begin{align*}
A(x) &= |\mathcal{R}(f(x))| = |\Re(f(x))|,
\end{align*}
\]

The local amplitude of the signal is give by \( A(x) = ||f(x)|| = \sqrt{f_x^2 + f_y^2} \), while its local orientation \( \theta \) and local phase \( \psi \) are specified by the following relations

\[
\begin{align*}
f = A \cos \psi, \quad f_1 = A \sin \psi \cos \theta, \quad f_2 = A \sin \psi \sin \theta.
\end{align*}
\]

From each response we extract the Autoregressive (AR) coefficients of a first order 2D model; energy and mean. For comparison purposes, we also extracted some common features referred in the literature, as shown in Table 1.

<table>
<thead>
<tr>
<th>Features</th>
<th>Acoustic attenuation coefficient, measured by the slope coefficient of the linear regression of intensities along the depth lines [4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-order statistics, including the mean, standard deviation and kurtosis of the pixel gray level;</td>
<td>Haar wavelet decomposition, where ( HH_{1,2} ) and ( HH_{1,2} ) are calculated the AR coefficients of a first order 2D model, ( (\alpha_{0,0}, a_{1,0}, a_{1,1}) ), energy and mean.</td>
</tr>
</tbody>
</table>

**Table 1: Features extracted from the US images**

Three feature sets are compared in the training phase: A) 29 Wavelet based features; B) 46 MF based features; and C) 99 features, representing all the extracted features. The selection of features, for each feature sets, was obtained by using the stepwise regression analysis [11] method (criteria to add: \( p < 0.05 \); to remove: \( p > 0.1 \)). This algorithm produced an optimal subset of 4 features for set A and 6 for set B and C.

The selected feature set is then fed into the Bayes classifier. The Bayes classifier assumes that the vector of features are multivariate normal distributed \([10, 15]\) with different means, \( \{\mu_1, \mu_2\} \) and covariance matrices, \( \{\Sigma_1, \Sigma_2\} \). The corresponding quadratic discriminant functions

\[
g_1(x) = \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) - \frac{1}{2} \log |\Sigma_1| + \log P(\epsilon_1),
\]

with \( \epsilon \in \{1, 2\} \), are used in the classification of a given feature vector \( x \) according with

\[
\begin{cases}
1 & \text{if } g_1(x) > g_2(x) \\
2 & \text{otherwise}
\end{cases}
\]

The performance of each feature set is analyzed by means of the overall accuracy (OA), sensitivity and specificity, in a leave-one-out cross-validation basis.

The Bayes factor is calculated to quantify the confidence level of the classification. It is defined as follows,

\[
\Lambda = g_1(x) / g_2(x)
\]

The values obtained for the normal and steatosis class are considered the baseline values and are used to assess the steatosis evidence in patients with CHC.

In the second phase, the test sample (CHC class) is used. The classification and Bayes factor for, each, is calculated and compared with the baseline values. The comparison is based on the mean and standard deviation of each class.
1.2 Experimental Results

A total of 106 US liver images from 106 patients were enrolled in the experiments: 36 patients constituted the control group; 38 were diagnosed with HS; and 32 with CHC. The patients were selected from the Gastroenterology department of the Santa Maria Hospital, in Lisbon, with known diagnosis based on liver biopsy results. The US images were acquired by expert operators in the hospital facility. A ROI of $128 \times 128$ pixels along the medial axis was extracted from each image.

The performances of the Bayes classifier with the 3 proposed feature sets are shown in Table 2. On comparison with feature set A, B set yields higher rate of correct classification. However feature set C outperforms feature set A and B. This set is composed of 2 features extracted from the wavelet transform, kurtosis and 3 features from the monogenic filter, as shown in Table 3.

### References


Detection of Children Speech Disfluency

Ángela Teixeira
http://www.uc.pt/fctuc/deec/

Carla Lopes
http://www.it.pt/

Fernando Perdigão
http://www.it.pt/

Department of Electrical and Computer Engineering,
University of Coimbra

Instituto de Telecomunicações - Polo de Coimbra
DEEC, Polo 2 of University of Coimbra

Abstract
Automatic Speech Recognition (ASR) can be applied to children language learning as a diagnostic tool for mispronunciations or speech therapy purposes. This is a well known challenge, not only because of the linguistic variability of children’s speech, when compared to adults’, but especially because of its acoustic characteristics. This study proposes an approach to detect mispronunciations or speech disfluencies in young children. The actual phoneme sequence and the likelihood values (provided by the Viterbi decoding of the speech signal) and the correct phone sequence are compared giving rise to an indication whether the word was mispronounced or not. Results show that the likelihood value may be used as a fairly good indicator of speech disfluency and other articulatory difficulties.

1 Introduction
Articulatory disorders affect a significant number of children especially in preschool age. It relates to a difficulty in the production of particular phonemes which are mispronounced, omitted or replaced. These disorders lead to verbal communication problems which get worse when the child begins its education, not only because of the difficulty in communicating but also due to the social problems that appear. Studies, conducted in several Portuguese regions, show that 20% to 30% of Portuguese children in preschool and first cycle ages require intervention at the level of speech therapy [8]. The differences in physiology of the vocal tract and the fast growth in children have long been the appointed causes for the lack of good results in the performance of speech recognition systems for children, particularly of young age [2] [3] [4]. Due to a shorter vocal tract, children have higher frequencies and higher fundamental frequencies (a 5-year-old child presents formant values 50% higher than an adult male [10]). Children speech also presents high spectral variability and speaking rate and high degree of spontaneity. These characteristics represent sensitive issues to develop effective acoustic, language and pronunciation models for reliable recognition of children’s speech.

In literature some of speech therapy systems can be found for several languages, namely English, [11], German, [12], Romanian, [13], Arabic, [17], Italian, [12], Spanish [16], Brazilian Portuguese, [14], etc. As far as we know, no European Portuguese system has yet been developed. In [14], it is presented a very interesting work for Brazilian Portuguese where several games were developed to improve phonarticulatory coordination of children and young deaf people (e.g., control of the airflow and fundamental frequency and positioning articulation of fricatives and vowels). Carvalho in [15] proposes an interactive application completely controlled by the utterance of 5 European Portuguese vowels.

Besides the lack of systems in Portuguese there is, to the best of our knowledge, no system that detects children’s speech disfluencies. This paper tackles this issue. A set of phone acoustic models has been developed in order to recognize the children’s pronounced phone sequence. Two grammars were used to decode this phone sequence: a grammar where the phone sequence is known and corresponds to transcription of the word the child is supposed to pronounce, and a free grammar with no restrictions in the phone sequence. The two phone sequences and the corresponding likelihoods were compared giving rise to an indication whether the word was mispronounced or not. Results show that the likelihood value may be used as a fairly good indicator of speech disfluency and other articulatory difficulties.

3 Disfluency Detection Procedure
To verify the accuracy of the word articulation, the values of the likelihoods calculated through the recognition process were analysed. Different grammars given to the recognition system, produced changes in the Viterbi paths and, as a consequence, different likelihood values.

Usually, for ASR systems, a list of words is given, typically represented by a network of word lattices (with the phone sequences permitted) to be used as a dictionary, with a list of which words the system can recognise. The recogniser can pick from those the one that best corresponds to the extracted parameters from the data being analysed. If the system knows the expected word a priori, it also holds the knowledge of the phone sequence and thus, it can force the recognition of that sequence of model phones. This corresponds to use a grammar (force alignment grammar) that gives the best alignment path between data and HMM models. If the phone models are robust, the likelihood calculated for each phone, and for the word as a whole, can be used as good indicators of weather the word was pronounced adequately or not.

In this study we propose that the system uses two different grammars during the recognition process: a force alignment grammar (the phone sequence is given to the recogniser) and a free grammar (the recogniser is free to choose the sequence of phones). Using a free grammar the Viterbi alignment is not conditioned to a grammar of words. At each step any phone can be chosen, resulting in a final sequence which may not correspond to a real word and thus presenting a lower likelihood. Figures 1 and 2 show examples where both grammars coincide in the sequence of phones (presented in the phonetic alphabet SAMPA) and where they differ. The times are indicated in units of cents of nanoseconds (10−9s). Figure 1a presents the alignment results using the force alignment grammar and b) using the free grammar. In this case, the two grammars produced the same results and even the same likelihood values. This generates strong evidence that the child uttered the word accurately. Considering the example in Figure 2, some changes can be observed in the outputs produced by the two grammars. The phone sequence attained with the free grammar (Figure 2b) differs from the expected sequence (force alignment grammar) having one suppression and one phone insertion. The correct word is “coroa” (kroa in Sampa) and the recogniser indicates that probably it was said “croa”. [kroa in Sampa].

2 Children phone recogniser
A children phone recogniser was developed using tools provided by the HKT toolkit [7]. Hidden Markov Model (HMM) acoustic models were built for each European Portuguese phone using HTK3.4. Each phone was modelled by a three-state left-to-right HMM. The input features were 12 Mel-frequency cepstrum coefficients (MFCC) plus $c_0$ (the zero-order MFCC coefficient), and their 1st and 2nd order time derivatives, computed at a rate of 10ms with a Hamming window of 15ms. The use of shorter windows than those regularly used in speech processing was chosen in an attempt to reduce the impact of the high pitch in the spectral parameters. The maximum likelihood criterion was used for training. Only acoustic models were employed in phone recognition: no language model was used. The phone models were trained by means of re-estimations and forced alignment. Additionally, the number of Gaussian Mixtures was increased up to 16 mixtures.

Both training and testing were carried out using a European Portuguese Children Speech Database for Computer Aided Speech Therapy [1]. It comprises 3726 audio files, collected from 111 children in kindergartens in the central region of Portugal, aged 5 to 6 years old.

In the audio recordings, the children would often hesitate or change their minds mid-sentence, laugh or breathe heavily and would rarely speak only the required word, thus resulting in several extra-linguistic and linking words in the annotations. For training purposes the original database files were cut, separating relevant speech words from the remainder of the audio. Both training and testing were carried out using this database and its original phone set [1]. The training set consisted of 80% of the data base and the test set consisted of the remaining 20%.

The model training process followed closely the instructions in [7] and a flat-start strategy was chosen to initiate the models since the label files had no phone-level information (only whole words were tagged). All the models were initiated equally, with mean and variance values equal to those of the global speech.
As expected, the recognizer which uses the free grammar produces likelihood higher values than the ones calculated forcing the Viterbi path through the expected phone sequence of the word. This has allowed us to infer about the quality of pronunciation of the word, i.e. about the detection of children speech disfluency.

4 Experiments and Results

Experiments where performed using the complete test data set. The recognizer ran a first time with a grammar allowing linking words, silence and non-linguistic events before a target word. This is necessary in order to spot a spoken word and its time limits. Then, using these limits, the recognizer is applied twice, using both the force alignment grammar and the free grammar. A discriminant measure is defined as the log likelihood difference between the two decoding results using the Viterbi algorithm. Finally, a decision about the correctness of the pronunciation is made. This information was summarised in the form of a binary classification problem with the hypothesis H0: ‘The child pronounced correctly the required word.’ Two errors can occur, a false rejection (FR) of a correctly pronounced word (type I error) or a false acceptance (FA) of a mispronunciation word (type II error).

In order to have a large number of mispronunciations to test the null hypothesis, we simulate a different expected word than the one uttered by the child randomly choosing a word from the list of existing words in the corpora for a given utterance. These values were used to understand how much the changes in the phone sequences would influence the likelihood values.

The trade-off between FA and FR errors are usually presented in a graphical form known as DET (Detection Error Trade-off) plot, which is shown in Figure 3 for our case. Each point in the DET plot represents a single chosen threshold. Using this graph, a good point may be to choose a threshold for which we get 5% of FA against 1% of FR.

References

Effect of Image Resolution on DLT-Lines Camera Calibration Including Radial Distortion

Manuel Silva
manuel.silva.thesis@gmail.com

Ricardo Ferreira, José Gaspar
ricardo.jag@isr.ulisboa.pt

Instituto Superior Técnico / UTL, Lisbon, Portugal
Institute for Systems and Robotics, Instituto Superior Técnico / UTL, Lisbon, Portugal

Abstract

Straight lines in 3D scenes and their images, allow to estimate the camera projection matrix and radial distortion, using the DLT-Lines calibration methodology. In this paper we detail the DLT-Lines and analyze the effect of image resolution on the calibration process, to help assessing, numerically, the benefits of improving the camera resolution.

1 Introduction

Laser range finders and color-depth (RGBD) cameras make the process of acquiring 3D data of a scene simpler. In particular, these sensors allow to obtain 3D lines/points on the scene that can be matched with lines/points imaged by a standard (RGBD) video camera. These matchings have been shown to be useful for the calibration of networked RGB cameras [4, 5].

A set of 3D-to-2D point correspondences and the Direct Linear Transformation (DLT) permit to estimate the camera intrinsic parameters, orientation and position in a global frame [2]. In case of having just image lines containing the 3D points, calibration is still possible using the so-called DLT-Lines methodology [5]. Considering the Division Model for radial distortion, introduced by Fitzgibbon [1], DLT-Lines also allows the calibration of cameras with radial distortion [5].

Given that DLT-Lines is based on 3D and 2D data (see example in Fig.1), it is prone to measurement errors both in the 3D and 2D data sources. In this paper we use the uncertainty caused by the 2D noise on the calibration process, to help assessing, numerically, the benefits of improving the camera resolution.

2 Camera Model and DLT-Lines

The perspective camera model describes the mapping of the 3D space to the 2D projective plane [2]. According to the pin-hole camera model, a scene point in homogeneous coordinates $M = [X Y Z 1]^T$ is imaged as a point $m = [u v 1]^T$:

$$m = PM = K[R t]M$$

where $\hat{m}$ denotes equal up to a scale factor, $P$ is a $3 \times 4$ projection matrix, $K$ is a $3 \times 3$ upper triangular matrix containing the intrinsic parameters of the camera, $R$ is a $3 \times 3$ rotation matrix representing the orientation of the camera and $t$ is a $3 \times 1$ vector representing the position of the camera.

The rotation and translation are defined with respect to a fixed absolute (world) coordinate frame.

The pin-hole parameters can be estimated using image lines and scene points [5]. The projection of a 3D line $L_i$ on the camera image plane can be represented by the cross product of two image points, in projective coordinates, $\hat{l_i} = m_{1i} \times m_{2i}$ [2]. Applying the multiplication by $L_i^T$ on both sides of Eq.1, leads to $l_i^T P M_{2i} = 0$ where $M_{2i}$ is a 3D point in projective coordinates lying in $L_i$. The properties of Kronecker product [3] allow to obtain a form factorizing the vectorized projection matrix:

$$(M_{1i}^T \otimes l_i^T) vec(P) = 0.$$  

(2)

Considering $N \geq 12$ pairs $(M_{2i}, l_i)$, one forms a matrix $B$, $N \times 12$, by stacking the $N$ matrices $M_{1i}^T \otimes l_i^T$. The DLT-Lines calibration data is illustrated in Fig.1. It consists of paired 3D points and 2D lines. Alternatively, given a 3D line $L_i$ and its projection represented by the image line $l_i$, any 3D point lying on the 3D line $L_i$ can be paired with 2D line $l_i$. On the other hand, any image line $l_i$ can be paired with any 3D point lying on $L_i$, i.e. more than one image line can be paired with a 3D point. The least squares solution, more precisely the minimizer of $\|B vec(P)\|^2$ subjected to $\|vec(P)\| = 1$, is the right singular vector corresponding to the least singular value of $B$.

Note that the pin-hole camera model, as presented on Eq.1, does not contain yet the radial distortion. To include radial distortion, we use Fitzgibbon’s Division Model. As proposed by Fitzgibbon [1] an undistorted image point, $\tilde{m}_i = [u_i v_i]^T$, is computed from a radially distorted image point $m_i = [u_i v_i]^T$ as $\tilde{m}_i = m_i/(1 + \|\tilde{m}_i\|^2)$, where $\lambda$ represents the radial distortion parameter. Fitzgibbon model allows to define a line $l_{12}$ as the cross product of two points:

$l_{12} = \begin{bmatrix} u_{1d} v_{1d} \\ v_{1d}^2 x_{1d}^2 + v_{1d}^2 y_{1d}^2 + v_{1d}^2 z_{1d}^2 + 1 + \lambda x_{1d}^2 y_{1d}^2 + \lambda y_{1d}^2 z_{1d}^2 + \lambda z_{1d}^2 x_{1d}^2 \\ 1 + \lambda x_{1d}^2 y_{1d}^2 + \lambda y_{1d}^2 z_{1d}^2 + \lambda z_{1d}^2 x_{1d}^2 \\ 1 + \lambda x_{1d}^2 y_{1d}^2 + \lambda y_{1d}^2 z_{1d}^2 + \lambda z_{1d}^2 x_{1d}^2 \end{bmatrix} = \hat{l}_{12} + \lambda e_{12}$

(3)

where $s_i$ is the norm of distorted image point $i$, $s_i^2 = u_{id}^2 + v_{id}^2$, the distorted image line is denoted as $l_{12} = [u_{id} v_{id}]^T \times [u_{id} v_{id}]^T$ and the distortion correction term $e_{12} = [v_{id} x_{id}^2, v_{id} y_{id}^2, u_{id} z_{id}^2, u_{id} x_{id} z_{id}^2, v_{id} x_{id} z_{id}^2, 0]^T$. Applying Eq.3 on Eq.2 leads to the following equation:

$$(M_{1i}^T \otimes (l_{12} + \lambda e_{12}) vec(P) = 0,$$  

(4)

which can be rewritten as

$$(B_{ki1} + \lambda B_{ki2}) vec(P) = 0,$$  

(5)

where $B_{ki1} = M_{ki1}^T \otimes l_{12}, B_{ki2} = M_{ki2}^T \otimes e_{12}$ and $M_{ki2}$ denotes the $k^{th}$ 3D point projecting to the distorted line $l_{12}$.

Considering $N \geq 12$ pairs $(M_{ki1}, l_{12})$, where $N = \max_i l_{max}$, one forms two $N \times 12$ matrices, $B_1$ and $B_2$, by stacking matrices $B_{ki1}$ and $B_{ki2}$. As proposed by Fitzgibbon, left-multiplying the stacked matrices by $B_1^T$ results in a Polynomial Eigenvalue Problem (PEP), which can be solved for example in Matlab using the polyeig function. It gives, simultaneously, the projection matrix, in the form of $vec(P)$, and the radial distortion parameter $\lambda$. 

Figure 1: DLT-Lines calibration methodology based on 2D-to-3D lines and points correspondences. The image data (a) consists of line segments (red) represented by points (blue stars), while the scene data is formed by 3D lines/points (b). Calibration involves estimating camera pose, $[R t]$, intrinsic parameters, $K$, and radial distortion, $\lambda$. 


Conclusions

The mean reprojection error, is effectively a good indicator of the accuracy of the calibration methodology. In terms of future work, we plan to further explore empirical and theoretical rules stating the required precisions of the 2D and 3D data in order to obtain a pre-specified precision of the results of DLT-Lines. These rules will serve the purpose of building user interfaces helping the in-situ calibration of networked cameras.

3 Experiments

In order to test the effect of different camera resolutions in DLT-Lines, we considered three synthetic cameras with standard resolutions and common 8mm lens in a 1/4 [in] CCD (intrinsic parameters). The cameras are considered to be at the world origin and aligned with the world frame (R = I).

All image points are distorted using Fitzgibbon’s radial distortion model, with λ ∈ {0, 10⁻⁷}. Figure 2 illustrates the experimental setup (a).

Considering a 3D line segment, L_i represented by two 3D points (see Fig. 2(a) blue dots), the corresponding image line l_i is defined as a collection of two or more points selected from the image line segment (red dots and lines, black circle and lines). The image line segment is computed using the camera projection matrix. Gaussian noise is added to the image points after computing the projected image line, and before obtaining multiple image lines, as cross products of pairs of distorted image points, which form the DLT-Lines input. The camera projection matrix and the distortion parameter are estimated using DLT-Lines (see Eq. 5), and the noisy and distorted image lines according to Eq. 4.

Figure 2 compares the performance of DLT-Lines. The comparison encompasses the mean reprojection error Fig. 2(b), the mean of the relative error of the estimated horizontal focal length Fig. 2(c) and, in the case λ = 10⁻⁷, the difference between λ and the estimated radial distortion parameter, λ_err. The plots involve 100 calibrations for each noise level (standard deviation, horizontal axis) and for each image resolution. Figures 2(b) and (c), show that in our setup the calibration errors of distorted and undistorted cameras are almost equal. In addition, the similarity of the mean reprojection error plot, Fig. 2(b) with K_err plot, Fig. 2(c) and λ_err plot, Fig. 2(d), shows that the common cost function used in nonlinear calibration, the mean reprojection error, is effectively a good indicator of the accuracy of the calibration methodology.

4 Conclusions

In this paper, we assess the benefit of image augmenting resolution in the precision of DLT-Lines. The assessment consists of testing the calibration methodology in a synthetic setup, encompassing a distorted or undistorted imaging system configured at different resolutions. The results show that the increase of image resolution improves the performance of DLT-Lines.

\[ \text{horiz. focal length relative error} = \frac{\lambda - \lambda \text{err}}{\lambda} \]

In terms of future work, we plan to further explore empirical and theoretical rules stating the required precisions of the 2D and 3D data in order to obtain a pre-specified precision of the results of DLT-Lines. These rules will serve the purpose of building user interfaces helping the in-situ calibration of networked cameras.

5 Acknowledgments

This work was supported by the FCT project PEst-OE / EII / LA0009 / 2011, by the FCT project PTDC / EECRO / 105413 / 2008 DCCAL, and by the project High Definition Analytics (HDA), QREN - I&D em Co-Promocão 13750.

References

Abstract
Iterative Closest Point (ICP) is an algorithm employed to minimize the difference between two Point Clouds. It is commonly used in medical imaging and computer vision applications in robotics, mainly for navigation tasks.

In this paper, ICP will be used to register 3D point clouds obtained from different scans of human bones, in order to localize a robot manipulator performing surgical tasks, e.g., drilling the femoral bone.

The Point Cloud Library (PCL) was used in this paper to register the 3D point clouds obtained from CT images. The results obtained from the registration algorithm showed a 5 mm Root Mean Square Error (RMSE), suitable for the on-going work.

1 Introduction
Currently medical/surgical methods aim to be less and less invasive. By using three-dimensional models, invasiveness can be reduced drastically with the help of various algorithms, allowing to accurately position surgical instruments at different locations of the human body, e.g., inside or outside.

This article focuses on orthopaedic surgery, where obtaining the position and orientation of the bones and surgical instruments is very important. The objective is to register the intra-operative position of the femoral bone to the pre-operative scenario. The pre-operative scenario is modelled using a 3D point cloud of the full bone, obtained from computer tomography (CT) medical images. The intra-operative scenario will be modelled using a 3D point cloud obtained from US images, i.e., a small part of the bone. It is also studied the best position of the US probe, to capture the intra-operative 3D point cloud, in order to obtain optimal results.

Medical image processing comprises a set of techniques used to extract information, helping healthcare professionals in their diagnoses. The medical imaging technology comprises a set of methods for data collection, ranging from the Ultrasound, (US) Computed Tomography (CT), Magnetic Resonance Imaging (MRI), among others.

Traditionally, these images are presented in the form of two-dimensional images obtained from devices using the international standard DICOM (Digital Image Communications in Medicine). In recent years have been developed software platforms capable of processing these two-dimensional medical images and get a 3D model of the structure under consideration, like ITK, the Segmentation & Registration Toolkit [1].

The 3D reconstruction of anatomical structures revolutionized medicine, allowing more detailed visualization of anatomical structures non-invasively.

2 Methods
Registration is the process of transforming different sets of data into one coordinate system. Image registration algorithms can be classified according to the transformation models they use to relate the target image space to the reference image space.

The first broad category of transformation models includes linear transformations, depicted in figure 1, which include rotation, scaling, and other affine transforms, but not translation (translation is a non-linear transformation) [2].

The second category of transformations allows ‘elastic’ or ‘nonrigid’ transformations. Non-rigid transformations include radial basis functions (thin-plate or surface splines, multi-quadrics, and compactly-supported transformations), physical continuum models (viscous fluids), and large deformation models (diffeomorphisms) [3].

One of the most used methods for 3D point cloud registration is the ICP (Iterative Closest Point) [4], that is an algorithm employed to minimize the difference between two clouds of points. ICP is also often used to reconstruct 2D or 3D surfaces from different scans, to localize robots and achieve optimal path planning, and to co-register bone models (which is the focus of the present paper).

The algorithm is conceptually simple and is commonly used near real-time. It iteratively revises the transformation (translation, rotation) needed to minimize the distance between the points of two raw scans, point clouds.

As inputs, the algorithm requires points from two raw point clouds, initial estimation of the transformation and a criterion to stop the iterative scheme.

The algorithm steps are:
- Associate points by the nearest neighbour criteria;
- Estimate transformation parameters using a mean square cost function;
- Transform the points using the estimated parameters;
- Iterate (re-associate the points and so on, i.e., until the stop criteria is met or the pre-defined max number of iterations is reached).

The Registration using the Iterative Closest Point (ICP) Algorithm from PCL [5], requires an input point cloud and a target point cloud. In this paper, the input will be obtained through several slices obtained with the aid of an ultrasonic probe, attached to a Polaris Spectra 3D tracking device. The target point cloud is obtained through a CT scan. The first step of the method is to determine pairs of corresponding points. After, an estimated transformation is obtained, that will minimize the distances between the correspondences. At the end of the iterative scheme, the transformation is applied to align/register the input and target 3D point clouds.

To obtain the error associated to the registration obtained it will be used the root mean square error (RMSE), depicted in equation 1.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (V_{g} - V_{\theta})^2}$$

where $V_{g}$ are the ground-truth values (obtained from the CT scan), $V_{\theta}$ are the values to be registered (obtained from a smoothed US scans), $n$ is the number of 3D points contained in the point clouds.

3 Experimental Results
The Point Cloud Library (PCL) is a large scale, open project [5] for 2D/3D image and point cloud processing. The PCL framework contains numerous state-of-the-art algorithms including filtering, feature estimation, surface reconstruction, registration, model fitting and segmentation. These algorithms can be used, for example, to filter outliers from noisy data, stitch 3D point clouds together, segment relevant parts of a scene, extract keypoints and compute descriptors to recognize objects in the world based on their geometric appearance, and create surfaces from point clouds and visualize them.

To obtain reliable results, of which the best method to use, it is necessary to keep in mind several main factors. The task that will be performed by the robot, the time available to perform the task, and the accuracy required. In the case of medicine the tasks executed are very
complex and usually they have to be performed as fast as possible and the probability of error should be null.

The first results obtained using the PCL based ICP, were performed with three point clouds. The first one obtained from the CT scan of the femur and representing the all bone, pre-operative scenario, NT, depicted in figure 2. The second (NC1) and third (NC2) point clouds represent two scans of the intra-operative scenario, i.e., the US scans. At this stage of the ongoing work, the two point clouds were obtained by: cropping the first point cloud and adding a known transformation (rotation of 10 degrees for each Euler angle and translation of 10 mm for each 3D direction). It was also added speckle noise to both point clouds, to simulate the US typical noise.

<table>
<thead>
<tr>
<th>Point Cloud 1</th>
<th>Point Cloud 2</th>
<th>RMSE [pixel]</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC1</td>
<td>NC1A</td>
<td>5.48665</td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>6.9609</td>
</tr>
<tr>
<td>NC2</td>
<td>NC2A</td>
<td>18.4754</td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>18.7774</td>
</tr>
</tbody>
</table>

Table 1 – ICP based registration results.

In figure 3 is depicted the result of the NC1 point cloud registration process, i.e., it was applied to it the transformation matrix obtained from the registration process. In table 1 is presented the RMSE error, 6.9609, that was obtained using the NT point cloud. If the user enters an approximate position of the point cloud, a crop of the NT cloud could be performed (obtaining NC1A) and the registration error will diminish as presented in table 1 and figure 4. From table 1, it can be observed that the 3D position of the point clouds relative to the bone point cloud, NT, is crucial to achieve good results, against the point clouds NC1A or NC2A.

The 5 mm error presented in table 1, shows good accuracy, although it is not suited for real time registration tasks. A first approach is to reduce the number of 3D points in the US cloud, lowering the computational burden of the registration. The algorithm used for this task creates a 3D grid as a set of small 3D boxes in space over the US point cloud data. Then, for each 3D box, its centroid is computed from all the points that belong to it. This approach assures a representative of the 3D points of the box, i.e., its centroid, and not the simple geometric centre of the 3D box. The down-sampling was adjusted with a 1.5 mm cube so that the object does not lose its original shape and keeping a small error of 1.5 mm in each direction.

In figure 5 it is clearly seen that after down-sampling the cloud have fewer points, making the cloud less dense than the cloud of the Figure 4. The results obtained with or without the down-sampling approach are similar in terms of error but, with it the processes is performed 20 times faster.

In the future it is expected that with the aid of a high quality GPU and an intelligent down-sampling method reduce severely the processing time.

4 Conclusions

In this paper was registered the intra-operative position of the femoral bone to the pre-operative scenario, using the ICP method from the point cloud library, with success. The pre-operative scenario was modelled using a 3D point cloud of the full bone, obtained from computer tomography (CT) medical images. The intra-operative scenario was modelled using a 3D point cloud simulating a 3D US scan, i.e., a small part of the bone. It was also observed that the 3D position of the 3D US point clouds relatively to the pre-operative point cloud is of particular importance to achieve good accuracy.

Future research will be focused in using real 3D US point cloud data to register the intra to the pre-operative scenarios, to investigate the optimal position of the 3D US point cloud, and to speed the registration task.

References

Shots Detection of Sport Headlines using the Angular Orientation Partition Descriptor

António M.G. Pinheiro
pinheiro@ubi.pt

Unidade de Deteção Remota
Universidade da Beira Interior
Covilhã, Portugal

Abstract
In this paper a method for detecting the limits of sport video sequences shots using the Angular Orientation Partition Descriptor, is described. This descriptor is particularly suitable for video processing because of its resilience to rotation and translation.

After the descriptor computation of each individual frame, the K-Means is computed, resulting clusters of frames. A final refinement of those frame clusters is achieved using the Euclidean distance between two adjacent frames descriptors and their statistical properties. The method is applied to football (soccer) headline sequences downloaded from youtube.

1 Introduction
Shot detection on video sequences might have multiple applications in several domains, like event detection or content extraction and analysis. Automatic summarization and browsing of the video content [2, 5] suitable to result in a faster access to video information, without the waste of computational and network resources, is of special importance. Eventually video parsing and associated labelling, iterative video applications, or video information retrieval and annotation based on content, are major applications where shot detection of video sequences might be very important. These are very important challenges for multimedia technology due to the large amounts of video information available, in particular over the internet, that demand efficient and reliable search functionalities. Nowadays, most of the existing content annotation is achieved through human annotations after visual inspection of the video sequences. With the available technology is usually difficult to navigate through the video sequences and find any specific information of interest. Recently a large number of research works on the domain of video segmentation [1, 2] or event detection on videos [4] have been published. The central idea behind this research is the development of methods that allow a high level description of the video information, based on its semantic content.

In this work, the Angular Orientation Partition (AOP) descriptor [3] is used for shot detection. The AOP provides reliable image description that allows very effective similar images recognition, resilient to image rotation and translation. These are the important characteristics that result in an effective shot detection over video sequences.

The main idea is to produce a reliable tool suitable for the temporal segmentation of a video sequence. The target application are low quality videos of Soccer headlines, as those usually available over the internet on online repositories (like YouTube). For testing, a set of YouTube videos with Soccer headlines was downloaded.

This introduction, is followed by the shot detection method description. Some results are shown. Finally, a conclusion will be reported.

2 Video Shot Detection
The video shot boundaries detection is based on the following steps:
1) Computation of the AOP descriptor for all frames of the video sequence.
2) Clustering the AOP frame descriptors using the K-Means algorithm.
3) Every isolated frame descriptor, located between two frames that belong to the same cluster, and that has a different cluster allocated, is changed to the neighbourhood one, to keep the cluster continuity. The same is done for every two joint clusters, located between two frames that belong to the same cluster.
4) Computation of the border regions using the Euclidean distance, $d_E$ between consecutive frames.
5) Mark all the extremes that have an Euclidean distance larger than a threshold distance $d_T$: $d_E > d_T = M_{10} + S_e \times \sigma_{10}$, where $M_{10}$ and $\sigma_{10}$ are the median and standard deviation of 10 samples of the Euclidean distance $d_E$ starting in the considered frame. $S_e$ can be considered as a sensibility factor, that was set to 2.5 experimentally with the objective of keeping a good balance between the number of False Positive and the False Negative. If $S_e$ grows the number of False Negative tends to increase. In opposition, if $S_e$ becomes lower the number False Positives tends to increase.
6) The shot limits are the frames that result in a cluster change and respect $d_E > d_T$ for that frame in a 2 frames neighbourhood (for both sides).

Clustering the AOP descriptors using K-Means result in a reliable detection of the different borders of the shot limits. In most of the cases K-Means clusters also result in very accurate segmentation of the shots. The first frame of a new shot will result typically in a new cluster. Unless two consecutive shots represent very similar scenes, K-Means will separate the different shots, grouping the AOP descriptors of the different frames in different clusters. In some cases, where the type of scene is very similar it might occur no separation. However, typical sport (football) headlines consist of sequences of shots with different zoom levels. The represented shot is usually repeated with a closer zoom, which makes the type of scene quite different. Typically these scene repetition is separated into different clusters. Even with two clusters separation (K=2) this tends to occur.

However, close shots (zooms in scenes) tend to have a high dynamic associated, that results in different clusters for the same shot. Moreover, long shots can result also in different clusters, because are represented by frame images with large temporal separation, that might have no similarity.

This extra clustering separation can easily be suppressed as shot boundaries, because usually consecutive frames in the same shot are very similar, and result in a small Euclidean distance between their AOP descriptors. Moreover, if two consecutive frames are in different shots, usually have a large Euclidean distance between them.

To improve the stability of the method, the median and the standard deviation of 10 consecutive frames were also computed, for the whole sequence. This standard deviation typically exhibits a peak in the proximity of every shot boundary. The standard deviation was multiplied by 2.5 (this value provides a good separation after observing many different tests) and summed with the median, resulting in a threshold distance.

A cluster change is only considered as a new shot if the Euclidean distance is larger than this value. This process usually avoids the separation in two of a long shot, because intermediate distances are small.

3 Results
The shot limits detection provided by this method results in typical Precision and Recall values of 90%. In some cases the Recall value becomes lower. However this value can be increased using a lower value of $S_e$. Examples of the Precision Recall values obtained for three youtube video sequences of soccer headlines is presented in table 1.

In figure 1 an example of shot detection in a video frame is given. In this figure, the black circles are locations where $d_E > d_T$, the * represent the real shots separation frames, and the triangles $\triangle$ the final shot limits. The blue line represents the different cluster variation on time, the dark cyan line represents the Euclidean distance between the AOP descriptor of consecutive frames, $d_E$, and the black line represents the threshold distance, $d_T$.

This threshold distance was also tested instead of the Euclidean distance $d_E$ for cluster separation validation, but was not so stable, and results in a higher number of False Negatives.

The value of K (of K-Means) was set to 5 because represents the best commitment between too many and only few initial segmentations in our experiments.

The represented test, results in 3 False Negatives and in 2 False Positives. This is rather good because in this case, the headlines video sequence is unusually large, with many zoom in (close distance) scenes,
that are usually more demanding. These scenes, are represented by regions with large unstability of the resulting clusters in figure 1.

The third false positive situation has proven to be the most difficult to deal with. This represents a typical situation in sports headlines video sequences, where the transition between two shots is made smoothly by superimposing the two shots during few frames.

The method was also tested with temporal sub-sampling and kept the same level of performance.

<table>
<thead>
<tr>
<th></th>
<th>Shots detected</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>22</td>
<td>90 %</td>
<td>86 %</td>
</tr>
<tr>
<td>Example 2</td>
<td>22</td>
<td>90 %</td>
<td>64 %</td>
</tr>
<tr>
<td>Example 3</td>
<td>22</td>
<td>91 %</td>
<td>91 %</td>
</tr>
</tbody>
</table>

Table 1: Precision Recall values for three examples of the youtube soccer headlines shot detection.

4 Conclusion

The AOP descriptor reveals to be very effective and resilient when used for segmentation of a video sequence.

Typically, the number of False decisions is kept low and a reliable model that can be used together with other methods was defined. In the performed tests, the percentage of True Positives was about 90%. Moreover, the number of False Positives is usually smaller than the False Negatives value, resulting in very high Precision.

For the near future, experiments using standard video sequences (like Trecvid sequences - have the problem of not being available to general public) will be carried out. This will allow an effective comparison between this shot detection method and some of the state of the art methods.

The reliability of the described method is also revealed for video sequences with high dynamics, as it is in the football (soccer) matches scenes used for testing.

That procedure also reveals to have a large resilience to different temporal sampling. When dealing with very low quality video, as it is in many cases the situation of videos downloaded over the internet the shot detection method reveals again a very reliable behaviour.

References


3D reconstruction of bat trajectories from stereo vision

Carlos Faneca¹
carlos.faneca@ua.pt
Ricardo Correia¹
rjiao@ua.pt
José Vieira¹
jnvieira@ua.pt
Carlos Bastos¹
cbastian@ua.pt
Carlos Fonseca²
cfonseca@ua.pt
Maria Pereira²
mariojoapereira@ua.pt
Miguel Mascarenhas³
miguel.mascarenhas@bio3.pt
Hugo Costa³
hugo.costa@bio3.pt
Joana Bernardino³
joana.bernardino@bio3.pt

¹Department of Electronics, Telecommunications and Informatics
University of Aveiro
Aveiro, PT
²Centre for Environmental and Marine Studies
University of Aveiro
Aveiro, PT
³Bio3
Lisbon, PT

Abstract
3D reconstruction of fast moving objects is an area with many potential applications. The goal of this paper is to evaluate the 3D trajectory of a flying bat in a controlled environment. Our solution consists in using a stereo vision system and compute an 3D position in each frame by selecting and matching a feature over rectified images.

1 Introduction
The purpose of this research is to develop a method to estimate the 3D trajectory of a flying bat in a large flight room. We used a stereo system with two monochrome cameras sensitive to infrared lights, for recording the flight trajectory of the bat. This trajectory was reconstructed by finding the bat in one image using background segmentation to select one point with the best features to be matched in the other image. The stereo system uses triangulation based on epipolar geometry to reconstruct the three-dimensional (3D) positions of the bat.

2 Equipment
The data acquisition was done in a large flight room with dimensions of 10×6×3 m; Walls and ceiling in the field of view were covered by white foam and white painted to get better contrast of the bat. The stereo vision system comprises two monochrome video cameras, two infrared illuminators and one processing computer. The cameras are Guppy PRO F-031B with maximum resolution of 640×480 running at 30 fps. Two sets of lenses with fixed focal length of 3.5 mm are employed, giving a field of view of 49°. As infrared illuminators we used two Raymax25 with a wavelength of 850 nm, an angle of 30° and 20 m of illumination range [4]. The cameras were placed next to each in a parallel configuration with a baseline of 350 mm. The illuminators were placed between the cameras to minimize the bat shadows. The stereo vision system was set in the bottom of the flight room (Figure 1).

Figure 1: Flight room.

3 Methods
Each experiment consists in the release of a bat in the flight room and recording the flight with both cameras for 35 seconds. The synchronization of both video files is achieved using a time stamp (flashing a LED on in the field of view of both cameras). Figure (2) provides an outline for the system that process the data acquired from the experiments.

3.1 Preprocessing
All computations are done offline, but a few processes and parameters are computed only once for all experiments since the position and the lighting of the stereo vision system is fixed. These processes and parameters are: calibration of the stereo rig, the evaluation of the projection matrices and the background estimation.

3.1.1 Calibration
The calibration of the stereo rig was done using the Camera Calibration Toolbox for Matlab[1]. A set of 20 images with different positions of a chessboard was simultaneously captured from each camera. The chessboard used had 7×8 squares with 45 mm between vertices. All the chessboard corners were extracted and refined to obtain the minimal error on both camera calibrations. This and the parameters used to rectify images were saved for later use.

Figure 3: Corrected geometry using the stereo rig calibration matrices.

3.1.2 Projection Matrix
The projection matrix [3] of both cameras was used to estimate the 3D coordinates. Eight points were manually extracted from the rectified background image from both cameras and the 3D coordinates of these points were measured (see Figure 3). These points were used to estimate the projection matrix for both cameras.
3.1.3 Background Estimation

Since the position of the stereo vision system and the lighting of the flight room are fixed in all experiments, the backgrounds were estimated using the median of several non-sequential rectified images from each camera.

3.2 Processing

The process consists of analyzing the video data from the stereo vision system and return the 3D position of the bat on each frame (see Figure 2). We start by rectifying each frame for both video files. The bat is detected on the left frame by a simple background subtraction. The result is a mask and it is used to select one point in the bat with the best features to be matched in the other image, this will be the darkest point on the bat. Since both images are rectified, the match of the point in the left image will be over the horizontal line with the same \( y \) coordinate in the right image (see Figure 4). At this point we have the image coordinate of a feature in the left image and the \( y \) coordinate of this feature in the right image. In order to find the \( x \) coordinate of this feature we subtract the row of pixels in the right image with the same row in the right background image. The value that we look for is the maximum difference of this subtraction. Since we have a match in the right image of the feature selected, we assume that the bat is visible on both frames and our next step is to refine this pair of matched points. This refinement is very important to obtain the minimum error on the 3D coordinate and consists in computing the maximum of a spline created by the difference between the row of pixels of the frame and the same row of the background image (see Figure 5). Finally we take this pair of points and reconstruct the 3D position of the bat by triangulation [2, 3]. This is performed by calculating the closest octagonal intersection of two 3D lines that crosses the pair of points. Each of these lines is computed through the projection matrix of each camera. This analysis is applied for all the frames and the result is the trajectory of the bat.

4 Analysis and Results

Applying this method to an experiment captured in 2012 we get the following results. The Figure (6) represents the estimated 3D trajectory of the bat. It is noteworthy that this data can be used to derive some measurements such as speeds and accelerations. The rate of detection can also be analyzed by manually counting the number of times that the bat is correctly detected in both frames. In this data we counted 240 occurrences and the system detected only 205, which gives a detection rate of 85%.

5 Conclusions

To effectively explore the flight behaviour of the bats, it is essential to obtain their 3D positions and trajectories. In this paper we presented a complete acquisition and processing system that reconstructs the 3D trajectory of a flying bat. The measurements of this work will be used to determine the effectiveness of an ultrasound bat deterrent.

References

1 Introduction

Sleep disorders form a class of medical conditions, pathological or not, affecting millions of people across the world. They are characterized by changes in the normal pattern of the circadian cycle and even sleep disruption, with severe consequences for the general health condition of the patient [2, 8].

A classification model and several temporal and position measures, are used to validate the results. It is shown that the movements recorded during each state are intrinsically different. They are characterized by the movement and information regarding the position of the patient. The hypothesis is that the movement and position of the patient interfere with his mobility and normal routines.

Table 1: Labeling rules: Each movement is labeled depending if the patient changed or maintained its position and if the sleep state changed or remained the same.

<table>
<thead>
<tr>
<th>Movement</th>
<th>Same Pos.</th>
<th>Change in Pos.</th>
<th>Same state</th>
<th>To/From Wake</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>To From Wake</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>To From Wake</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>To From Wake</td>
</tr>
</tbody>
</table>

2 Methods

2.1 Data pre-processing

A simple movement detector was implemented to detect movements and normal routines are easily detected, an accurate detection of the sleep and wakefulness states is still an open issue. The movement and position of the patient interfere with his mobility and normal routines.

2.2 Feature extraction

Each detected movement is labeled from 1 to 8 according to the following rules:

- If the patient changed or maintained its position and if the sleep state changed or remained the same, the movement is labeled as:
  - 1: Same state, Same Pos.
  - 2: Change in state, Same Pos.
  - 3: Same state, Change in Pos.
  - 4: To/From Wake, Same Pos.
  - 5: To/From Wake, Change in Pos.
  - 6: Same state, To/From Wake
  - 7: Change in state, To/From Wake
  - 8: To/From Wake, To/From Wake

For instance, if the patient changed position and the sleep state remained the same, the movement is labeled as 4: To/From Wake, Same Pos. This is done to ensure that the algorithm correctly identifies the state changes.

The Act data was collected with a Somnomedics device, placed in the non-dominant wrist of the subjects with sampling rate of 1Hz. The output is used to identify the sleep and wakefulness states and position. The Act data is used to validate the results. It is shown that the movements recorded during each state are intrinsically different. They are characterized by the movement and position of the patient. The hypothesis is that the movement and position of the patient interfere with his mobility and normal routines.

The detection of simple trends, such as circadian rhythm shifts, is also used in the estimation of sleep and wakefulness states. Some use activity count thresholds [2, 5] and others estimate the state by using statistical approaches [8].

In this work, the data pre-processing is presented, followed by the detection of movements and normal routines.

3 Results

It is shown that the movements recorded during each state are intrinsically different. They are characterized by the movement and position of the patient. The hypothesis is that the movement and position of the patient interfere with his mobility and normal routines.

The Act data was collected with a Somnomedics device placed in the non-dominant wrist of the subjects with sampling rate of 1Hz. The output is used to identify the sleep and wakefulness states and position. The Act data is used to validate the results. It is shown that the movements recorded during each state are intrinsically different. They are characterized by the movement and position of the patient. The hypothesis is that the movement and position of the patient interfere with his mobility and normal routines.

References


The subset yielding the highest accuracy is finally combined and three voting policies are tested, the obtained Sensibility, Specificity and Accuracy are shown in Table 3.

<table>
<thead>
<tr>
<th>Voting policy</th>
<th>Sensibility</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product</td>
<td>76%</td>
<td>68%</td>
<td>73%</td>
</tr>
<tr>
<td>Median</td>
<td>68%</td>
<td>67%</td>
<td>63%</td>
</tr>
<tr>
<td>Vote</td>
<td>65%</td>
<td>60%</td>
<td>64%</td>
</tr>
</tbody>
</table>

The Product combiner of the set of classifiers yields the highest accuracy, 73%, sensibility, 76% and specificity 68%, followed by the Median combiner. The fusion of 4 expert classifiers increases, not only the overall accuracy, but also the specificity to acceptable values.

### 4 Conclusion

In this paper a classification framework is presented to discriminate between movements recorded on nocturnal Actigraphy during Sleep and Wakefulness states. While the obtained accuracy, 73%, is not enough for an accurate estimation of the state along the entire night, the obtained results clearly show that movements that are apparently similar can in fact contain valuable information regarding the state of the patient.

Future work will focus on the addition of new features to the framework, such as HRV and Breathing based features, to further increase the classification accuracy.

### References


Table 2: Accuracy obtained with 3 different classifiers for the 4 subsets of features.

<table>
<thead>
<tr>
<th>Subset</th>
<th>LDC</th>
<th>SVM</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64%</td>
<td>65%</td>
<td>62%</td>
</tr>
<tr>
<td>2</td>
<td>68%</td>
<td>67%</td>
<td>63%</td>
</tr>
<tr>
<td>3</td>
<td>65%</td>
<td>60%</td>
<td>64%</td>
</tr>
<tr>
<td>4</td>
<td>56%</td>
<td>55%</td>
<td>54%</td>
</tr>
</tbody>
</table>

Table 3: Sensibility, Specificity and Accuracy obtained for the Product, Median and Vote classifier combiners.

2.3 Classification

The classification procedure aims at discriminating the movements observed during Wakefulness and Sleep state. The set of features extracted from movements with labels [1, 3, 5, 7] is labeled as class Wakefulness, on the other hand, features extracted from movements with labels [2, 4, 6, 8] are labeled as class Sleep.

The complete set of features is then divided into 4 subsets, i) a subset with length and energy of movement (Subset 1), ii) a subset with the AR coefficients and AR prediction error (Subset 2), iii) a subset with the parameter of the exponential distribution (Subset 3) and finally, iv) the discretized position (Subset 4).

Each subset of features is used to train i) a linear discriminant classifier (LDC), ii) a support vector classifier (SVC) with a linear kernel and a k-nearest neighbour (KNN) (3 neighbours).

The accuracy of each classifier is computed using a k-fold (k = 20) leave-one-out crossvalidation and the classifier yielding the highest accuracy is selected as the expert for each subset of features.

A fusion of classifiers is finally performed, testing several voting policies.

### 3 Results

Table 3 resumes the results obtained for the individual subsets of features. The subset yielding the highest accuracy is the second subset (features extracted from the AR model) using a LDC, followed by the first and third subsets using a SVC and an LDC respectively.

While the obtained accuracies are acceptable, the obtained specificities are in the order of 55% for all the considered classifiers and subsets, meaning that the classifiers tend to detect Wakefulness when the true state is sleep.

---

**Figure 2:** a) Actigraphy data (top), b) Hypnogram (middle) and c) Position (bottom). Detected movements are marked as black dots.

- Length of the movement, \( N \).
- Energy of the movement, given by \( E = \sum_{i=1}^{N} m_j(i) \).
- AR model (order 4) coefficients \( \theta \), given by the Yule-Walker equations.
- AR model prediction error.
- The parameter of Exponential distribution fitted to the data, given by \( f = \frac{1}{\mu_i} \).
- Discretized position (value from 1 to 5)
Abstract

Heart rate variability (HRV) analysis is the preferred tool to monitor the autonomic nervous system (ANS) activity, since the heart is its main effector organ. However, the cardiac variability strongly depends on the sympathetic and parasympathetic coupling dynamics with the sinoatrial node (SAN) and with the mechanical and physiological parameters of the heart itself. It is thus fundamental to describe the bias introduced by the SAN response to vagal and sympathetic stimuli and its influence on the computation of the HRV.

In this paper a physiological based incremental linear model, describing the dynamics of the SAN innervation is proposed aiming at accurately estimate the ANS activity by HRV analysis. Experiments with real data are used to validate and tune the model, confirming the results in the literature. ¹

1 Introduction

Heart rate (HR) is controlled by impulses originated in the SAN. The SAN is 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. The ionic mechanism underlying this pacemaker activity is a cyclic activation of the funny current (I_f) [5, 6].

This hyperpolarization-activated current main characteristics are: i) it is conducted by hyperpolarization-activated cyclic nucleotide-gated (HCN) channels [1], ii) its activation has a voltage dependency on the HCN channel which iii) are permeable to Na⁺ and K⁺ ions, and iv) it is highly dependent on cyclic adenosine monophosphate (cAMP)[10].

The funny current increases the membrane potential which causes the activation of Ca²⁺ channels (T-type and L-type)[6, 9, 13]. When the membrane voltage reaches the threshold, a new action potential is initiated triggering a new heart cycle.

The SAN has an intrinsic discharging rate of about 100 beats per minute, but in order to keep homeostasis, i.e. adapt the HR according to the changing needs, the SAN is innervated by the two branches, sympathetic and parasympathetic, of the ANS [9].

Stimulation of the sympathetic nerves releases norepinephrine (NE), which increases the rate and force of contraction of the heart by changing the HCN channels properties and the cAMP level[6, 9, 13]. The activation of beta-adrenergic receptors stimulates protein kinase A (PKA)[8] and increases cAMP concentration [10]. By this, the sodium-calcium pathway permeability is increased thus increasing ICa and I_f [1, 2, 10]. SAN myocytes become more excitable and the action potential rhythm rises.

Parasympathetic stimulation decreases the rhythm of sinus depolarization [11]. The acetylcholine (Ach) released at the muscarinic receptor increases the conductance of the potassium channels. The hyperpolarization of the SAN myocytes appears to involve direct gating of a K⁺ channel by a G-protein activated by the M2 muscarinic receptor[1]. Furthermore, cAMP level and the L-type Ca²⁺ current decreases [7]. As a result, the membrane resting potential decreases from −65 to −75 mV in response to vagal stimulation, reducing sinus rhythm and thereby decreasing the HR [3, 6, 9, 13].

In this paper an incremental model for the dynamics of the SAN and its dependency on the ANS is proposed. The model is formulated taking into account the system’s physiological characteristics. The model is validated with results found in the literature and with real data, acquired from 5 healthy subjects.

2 Problem Formulation

In this section, the SAN is described by an incremental linear model. Here, the dynamics of the parasympathetic and sympathetic components of the SAN are modeled by two IIR low pass filters. The filter’s transfer functions mimic the dynamics of the parasympathetic and sympathetic systems, these stimuli suffer different time delays. Thus, the release of Ach and NE in the SAN is affected by a time lag. The parasympathetic stimuli take 210msec to arrive to the SAN [7] while the sympathetic stimuli take 950ms [7]. These time delays are not considered in the described model but are contemplated in the generation of the initial signal, P(t) and, S(t). Each filter input ˜P(t) and ˜S(t) has a constant and an incremental component expressed as

\[ \dot{P}(t) = P(t) - P_0, \quad P_0 = 0.4 \]

\[ \dot{S}(t) = S(t) - S_0, \quad S_0 = 0 \]

where \( P_0 \) and \( S_0 \) are the cardiac vagal and sympathetic tones respectively.

The dynamics of the Ach and NE in the SAN are modeled by low pass filters denoted in Figure 1 by i) and ii). The two filters coefficients were chosen taking into account that the maximum rhythm changes is observed 0.63sec and 4.6sec after parasympathetic and sympathetic stimulation, respectively, as shown in Figure 3. The coefficients were thus set to \( T_1 = 0.063 \), \( T_2 = 0.46 \). The inputs are affected by a time delay that corresponds to the diffusion of neurotransmitters and to the metabolic modification latency.

The gain in filters i) and ii), represented by a and b, is unitary. The final parasympathetic gain is positive thus increasing the SAN depolarization period \( \dot{T} \). On the other hand, sympathetic stimulation decreases \( \dot{T} \) and the final gain is negative.

In block iii) the two signals are added and form \( \dot{T} \) which adds to \( T_0 \), the intrinsic SAN rhythm. Thereby, the model output, \( T(t) \), contains a basal rhythm and the ANS influence.

3 Experimental Results

The ANS activity is simulated using the model described in the previous section. The simulation is performed with a constant vagal tone and two impulses, a sympathetic and a parasympathetic, are generated.

The two impulses and the model response are shown in Figure 4. This simulation mimics the time delay, maximum sympathetic and parasympathetic effects on the depolarization period and the time of 90% recovery of
the initial value. The obtained results are in accordance with the responses obtained from animal experiences by Spear et al., shown in Figures 2 and 3. In this simulation the recovery of the initial depolarization rhythm is faster than in real data experiments which suggests that the metabolic processes in SAN are non-linear.

Figure 4: SAN node depolarization period after sympathetic and parasympathetic stimuli, results of model simulation

In order to test the proposed model with real data, RR data [12] was acquired from 5 healthy subjects (3 women and 2 men). Each subject performed an orthostatic test inducing a sympathetic activation in the ANS. The test was repeated 2 or 3 times, resulting in 13 datasets. All the data was acquired using the system described in [4].

Data processing consists in finding the maximum RR interval variation following the stimulus (induction of orthostatic position of body), and computing the time delay between the two points. The orthostatic test was chosen as this is the standard procedure when stimulating sympathetic activity. The performed tests do not include parasympathetical activity since this activity cannot be measured with non-invasive methods.

The results obtained for the maximum sympathetic activation using RR data, shown in Table 1, are in agreement with the results predicted by Spear et al. [7]. The main differences, not shown in Table 1, lie in the recovery time after the maximum sympathetic activation. This difference is expected since each subject maintains the orthostatic position after the test, thus maintaining sympathetic stimulation.

Table 1: Mean time of Maximum Sympathetic Activation and standard deviation for the HR measurements during the Orthostatic Test

<table>
<thead>
<tr>
<th>Events</th>
<th>Delay (sec)</th>
<th>Max. Activation(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>undetected</td>
<td>8.3804 ± 2.9988</td>
</tr>
</tbody>
</table>

Nevertheless, the level of agreement between the results suggests that the human SNA dynamics can be approximated by a linear model.

4 Conclusions

In this paper the intrinsic SAN metabolic dynamics of Ach and NE are characterized and their individual contributions are included in a model describing the SAN dependency on the ANS.

The obtained results show an agreement between the simulations and the responses of an animal SAN to vagal and sympathetic stimuli. While the maximum response time is virtually identical, the recovery dynamics are slightly different, suggesting that a non-linear model might yield better results.

The results using RR data show that the human SAN response to a sympathetic stimuli can be roughly comparable to the SAN of rabbits. The differences lie in the different activation levels and maximum response times, possibly due to physiological differences between species and errors in RR acquisition.

It was thus verified that the SAN activity and HR variations due to ANS stimuli can be represented and simulated by a complex system with non-trivial response. Further work will try to adapt this linear model to human physiology and study the mechanism of SAN normal rhythm recovery after ANS stimuli.

References

Towards the Evaluation of a Fuzzy Logic Model to Describe SBP and RR Relation: a Case Study Report

Sónia Gouveia
sonia.gouveia@ua.pt
Susana Brás
susana.bras@ua.pt

Abstract

This work presents a case report on the evaluation of a fuzzy logic model to describe the relation between the tachogram (RR) and the systolic blood pressure (SBP) values of a healthy subject. From the real data, both random/surrogate realizations were generated, in order to destroy/keep the original SBP and RR relation, respectively. The results point out that the fuzzy logic model is able to capture more information than simply the mean RR values, in particular, at SBP values at the left and at the right of the SBP median. These results are corroborated by the location of the membership functions over the simulations: for random data, the cluster centers are frequently positioned in the SBP median while, for surrogate data, the clusters are distributed around the SBP values contributing with extra information, at the left and at the right of the SBP median.

1 Introduction

Over the past years, the quantification of arterial-cardiac baroreflex sensitivity (BRS) has been useful in the study of many pathological states, where lower BRS have been associated with increased cardiovascular disease-related mortality [5]. Spontaneous time domain methods for BRS estimation typically assume linearity between SBP and RR values and the BRS is taken as a slope between the SBP and RR values identified in baroreflex related segments [3]. Therefore, local linearity is assumed in a segment and the BRS index is not a function of the SBP value.

These assumptions can be in contradiction with physiology, in particular, if drug induced methods are used for BRS estimation. In comparison with the spontaneous ones, drug induced techniques stimulate a larger and clearer SBP change in order to force a pronounced RR response (i.e., clearer baroreflex activation). Therefore, these methods explore the baroreflex function over a wider SBP range, while spontaneous methods allow the BRS assessment near the subject’s operating point. Invasive BRS methods assume a sigmoidal SBP–RR relationship and, therefore, allow for the BRS index to vary according to the SBP value.

In this work, we seek to answer “Should the BRS index be a function of the SBP value also for spontaneous data?". In a recent work, we explored the use of fuzzy logic models to describe the relation between SBP and RR values [2]. Fuzzy logic models do not assume a shape for SBP and RR relation and open the possibility to model non linear SBP and RR relation, which ultimately will make possible to obtain a BRS index as a function of the SBP value. Based on the EuroBaVar database with recordings of several of subjects [6], it was found that fuzzy models estimated from real data exhibited a significantly lower average modeling error than that estimated from random data, what motivated this study.

Here, an evaluation of a fuzzy logic model obtained for a healthy subject is presented. The evaluation was carried out by comparing the real fuzzy model with those estimated from random and surrogate realizations of the real data. The random data is obtained by scrambling and, therefore, keeps the probability density function of the real data while destroying temporal dependencies and SBP/RR relation. The surrogate data was generated by maximum entropy bootstrap, so keeping the main features and associations of the data.

2 Fuzzy Logic Model Estimation

The fuzzy description is based on If Then rules. Basically, the system input is mapped into membership functions that associate an input value to a membership degree. This value is then used to obtain the rule weight and the rule output is used to generate one consequent (the Then).

In this work, the fuzzy logic system was defined as a Sugeno model, which considers the system output as a function $z = f(x)$, i.e. $\text{RR}=f(\text{SBP})$.

Given the input $x$, a typical rule $i = 1, 2, ..., N$ with output $z_i$ is defined as

$$f_i(x)\mathbf{w_i},$$

where $f_i(x)$ are fuzzy sets and $a_i$ and $c_i$ are constants [4]. Each rule output $z_i$ is then weighted by its firing strength $w_i = \Gamma_i(x)$, where $\Gamma_i$ is a Gaussian membership function defined by its center $\mu_i$ and standard deviation $\sigma_i$ [1]. The final output of the system $\hat{z}$ is the weighted average of all $z_i$.

The number of rules $N$ and the parameters $a$, $c$, $\mu$ and $\sigma$ for each rule $i = 1, 2, ..., N$ were estimated by ANFIS [4]. The initial values for $\mu$ and $\sigma$ were obtained using subtractive clustering. The first center $(\mu_i)$ is identified as the point with maximum likelihood, i.e., the median of $x$ [1]. The next center is estimated as the previous, disregarding the data already assigned to existing clusters. The procedure stops when all data falls within a cluster [1]. The method, iteratively, divides the antecedent domain into clusters, estimating their centers [1], based on a predefined radius (cluster influence within the data space). The membership functions appear as the projection of these clusters on the $x$ axis.

Figure 1 presents an illustrative application of the fuzzy logic method to a real set of SBP and RR values. Figure 1(a) presents the real data superimposing the generated system output, i.e. the fuzzy surface describing the RR as a function of SBP values. The fuzzy surface is obtained from the membership functions represented in Fig. 1(b).

Figure 1: (a) Dispersion diagram of the first 512 SBP and RR values from the EuroBaVar datatfile A001LB, superimposing the estimated fuzzy surface $\hat{z}$. (b) Membership degree $\Gamma_i$ $(i = 1, 2, ..., 6)$ associated to each membership function, according to the SBP input values.

3 Experimental Data

The real data used in this work consists of one recording from the EuroBaVar dataset [6]. The A001LB file includes beat-to-beat tachogram (RR) and systolic blood pressure (SBP) series extracted from simultaneous spontaneous ECG and arterial blood pressure (ABP) signals, respectively. The records are of approximately 10 minutes length and were acquired at 500 Hz of sampling frequency. The fuzzy methods were applied to the first 512 beats of the SBP and RR series, considering each SBP value aligned with the following RR value, in accordance with previous studies [3]. More specific details on protocol procedure can be found in [3, 6]. From A001LB file, 1000 random replicas and 1000 surrogate replicas of the original RR series were generated while maintaining the original SBP

117
series. By one hand, the random RR replicas were generated by scramble the original RR values to produce a series with a random order [7]. By the other hand, the surrogate RR replicas were generated by maximum entropy bootstrap, thus preserving the temporal structure and the non stationary behavior of the original RR series in the surrogates [8]. For illustration purposes, Fig. 2 presents an excerpt of the original RR series, superimposing a random and a surrogate realization. It can be observed that the scramble of the original RR series produced a random realization with destroyed temporal dependency and, consequently, with no relation with the original SBP series. On the contrary, surrogate series kept the temporal structure of the original RR series and therefore held the main links with the original SBP series.

Figure 2: Excerpt of the original RR from A001LB datafile (red) superimposing a random (gray) and a surrogate realization (black).

### 4 Results

Figure 3(a) shows the 1000 fuzzy surfaces obtained for the random realizations (in gray) as well as the fuzzy surface obtained for the real data (in red). It can be observed that gray/random surfaces are more flatter than the red/real surface, evidencing that fuzzy surfaces on random data explain typically the mean RR value and constitute a null model. Therefore, the gray area composed by the 1000 random fuzzy surfaces constitute a band of possible variations for the surfaces assuming that there is no relation between SBP and RR. In this context, it can be interpreted that the real fuzzy surface has amplitudes different from those of the random fuzzy surfaces, in particular around the locations of the membership functions.

Figure 3(b) shows the estimated pdf’s of the locations for the membership functions over all realizations, distinguishing random (gray) and surrogate (black) realizations. Dashed lines localize mean SBP and mean RR values.

Figure 3: (a) Fuzzy surface obtained for the real data (red) and for the random realizations (gray); (b) Estimated probability density functions (pdf) of the chosen locations \( \mu \) for the random (gray) and surrogate (black) realizations. Dashed lines localize mean SBP and mean RR values.

Table 1: Number of fuzzy logic membership functions (# MFs) optimized in each one of the random and surrogate realizations.

<table>
<thead>
<tr>
<th></th>
<th>random</th>
<th>surrogate</th>
</tr>
</thead>
<tbody>
<tr>
<td># MFs</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

### 5 Conclusions

This case report on a healthy subject evidences that fuzzy logic to model SBP and RR relation can capture singularities, besides median values. These singularities are located in ranges with lower data density which, being distinct from that obtained from random data, are expected to provide additional and useful information concerning SBP and RR relation.

### 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 project FCOMP-01-0124-FEDER-022682, FCT ref. PEst-C/EEI/UI0127/2011, Instituto de Engenharia Electrónica e Telemática de Aveiro, IEETA/UA, Aveiro, www.ieeta.pt.

1. Gouveia is with IEETA/UA and acknowledges the postdoctoral grant by IEETA/UA (ref. BPD/UI62/24870/2012), sonia.gouveia at ua.pt
2. Brás is with IEETA/UA and acknowledges the postdoctoral grants by IEETA/UA (ref. BPD/UI62/5259/2012) and by the FCT project "Vital Responder" (ref. CMU-PT/CP0/0046/2008, funded by "Carnegie-Mellon|Portugal" program), susana.bras at ua.pt

### References


Predicting Malignancy from Mammography Findings and Surgical Biopsies

Pedro Ferreira¹
pedroferreira@dcc.fc.up.pt
Nuno A. Fonseca¹
nfi@ebi.ac.uk
Inês Dutra²
ines@dcc.fc.up.pt
Ryan Woods³
rwoods@gmail.com
Elizabeth Burnside³
EBurnside@uwhealth.org

¹CRACS INESC-TEC LA
University of Porto, Porto, Portugal
²Department of Computer Science & CRACS INESC-TEC LA
Faculty of Sciences
University of Porto, Porto, Portugal
³Department of Radiology, Johns Hopkins Hospital, Baltimore, MD, USA

1 Introduction

We applied machine learning methods to 348 consecutive breast masses that underwent image guided or surgical biopsy performed between October 2005 and December 2007 on 328 female subjects. These 348 findings are defined by 14 attributes, with one of them indicating if the finding is malignant or benign. Our main objective is to produce models that can have a good performance at predicting malignancy and a good performance at avoiding to expose healthy women to extra surgical or screening procedures. We are also interested in studying the actual relevance of mass density in the findings, since this is one of the attributes that usually is not regarded relevant by physicians. According to physicians, mass density is a feature usually considered to be difficult to annotate, because of the breast tissue, and fat composition. Previous works have shown that mass density can be an important attribute when predicting malignancy [3, 10, 11]. The 348 mammographies used in this study have annotations of mass density, which allow to (1) investigate in more detail the role played by this feature, and (2) produce models to predict this particular feature and help physicians distinguish between high and iso/low densities.

Much work has been done on applying machine learning techniques to study breast cancer, one of the most common kinds of cancer in the world. Our work, rather than focusing on the images, focuses on the medical reports. More details about the experiments, parameter settings, more references and explanations can be found elsewhere [4, 5].

2 Methodology

We use the same data set used by Woods and Burnside [11]. This data set is unique in the sense that all findings were retrospectively assessed and all of them have accurate information about the density of the breast masses. In that work, they showed that high breast mass density is a significant predictor of malignancy, even after controlling for other well-known predictors of malignancy such as mass margin and mass shape. The metric used to evaluate performance was interobserver agreement and they found a moderate k-value for mass density (0.53).

Each one of the 348 cases refers to a breast nodule retrospectively classified according to the Breast Imaging Reporting and Data System (BI-RADS) created by the American College of Radiology. On the other hand, a clinical radiologist assessed (at the time of imaging and without biopsy results) the density of 180 of these masses, in an evaluation that can be considered as “performed under stress” (prospective assessment). Pathology result at biopsy was the study endpoint.

When learning models to predict malignancy the attribute outcome is the target class. It assumes values malignant and benign and was determined using the results of biopsies. From the 348 cases, 118 are malignant (\(\approx 34\%\)), and 84 cases have high mass density (\(\approx 24\%\)) retrospectively assessed. Other attributes are mass shape, mass margins, depth, size, among others. For the purpose of our study, we have two attributes that represent the same characteristics of the finding, but with different interpretations. These are retro_density and density_num. Both represent mass densities that can assume values high or iso/low. Retro_density was retrospectively assessed while density_num was prospectively (at the time of imaging) assessed. These two attributes are our target classes when learning models to predict mass density.

The whole data set was split into two subsets: (1) training set: 180 cases, whose mass densities (density_num) were classified by a radiologist and (2) test set: 168 cases, whose mass densities (retro_density) were not annotated at the time of imaging, but instead in a reassessment of all the 348 exams performed by a group of experienced physicians.

Five experiments were performed with 180 findings (training set) while the remaining were performed with 168 findings (test set). From the five, the first three predict outcome and the other two predict mass density. The experiments can be described as follows. \(E_1\) predicting outcome using the attribute retro_density. \(E_2\) predicting outcome using the attribute density_num. \(E_3\) predicting outcome without any mass density information. \(E_4\) predicting mass density based on retro_density. \(E_5\) predicting mass density based on density_num.

We evaluated several classification algorithms available in WEKA [6] and varied their parameters. The experiments were performed with the WEKA’s Experimenter module using 10 times 10-fold cross-validation on the training dataset. For each algorithm we selected the combination of parameters that produced the best classifiers, and then selected the top three classifiers for generating models: NaiveBayes [7], DTNB (a decision table algorithm whose leaves are Bayesians networks) and SMO (a support vector machine [9] implementation [8]). A fourth classifier was selected, J48 (decision tree based on Quinlan’s C4.5 algorithm), due to its ability to produce readable and easily understandable models.

The last six experiments apply the models generated by the first five experiments (\(M_1, M_2, M_3, M_4, M_5\) and \(M_6\)) to the test set containing 168 cases as follows. \(E_6\) generates the values for mass density using the model trained with the attribute retro_density as the class variable (obtained by experiment \(E_3\)). \(E_7\) generates the values for mass density using the model trained with the attribute density_num as the class variable (obtained by experiment \(E_5\)). \(E_8\) predicts outcome using the model \(M_1\) trained with the attribute retro_density (obtained by experiment \(E_1\)), and uses the actual values of the attribute retro_density available in the test set. \(E_9\) predicts outcome using the model \(M_2\) trained with the attribute retro_density (obtained by experiment \(E_1\)), and uses the mass density values filled up by experiment \(E_6\) in the test set. \(E_{10}\) predicts outcome using the model \(M_3\) trained with the attribute density_num (obtained by experiment \(E_2\)) and uses the mass density values filled up by experiment \(E_7\) in the test set. \(E_{11}\) predicts outcome with the model \(M_4\)
that does not use any information about mass density, obtained in experiment E3. For this experiment, no mass density attribute is used in the test set.

We used the metrics CCI (Correctly Classified Instances, a.k.a. accuracy), F-measure (harmonic mean between Precision and Recall) and Kappa statistics to assess the classifiers. Whenever applicable we performed significance tests using paired t-test ($\alpha = 0.05$).

3 Results

Training to predict outcome In experiments ($E_1$, $E_2$) and ($E_3$), the best classifiers found were based on SMO. First of all, these results show that mass density has some influence on the outcome, specially when mass density is the one observed on the retrospective data (experiment $E_1$). The results obtained with experiments ($E_2$), ($E_3$) and ($E_4$) confirm findings in the literature regarding the relevance of mass density [1, 2, 10, 11], and also show that good classifiers can be obtained to predict outcome (with a high percentage of correctly classified instances and good values of precision and recall, according to F).

Training to predict mass density Our set of experiments $E_4$ and $E_5$ are related to predicting mass density. The best classifier for predicting retro_density was SMO and the best to predict density_num was NaiveBayes.

During the prospective study, the radiologist predicted 70% of masses on the 180 findings compared with the annotated masses of the retrospective study. The SMO classifier predicted 81.3% of correct instances when training on prospective data, and the NaiveBayes predicted 67.2% of correct instances when training on prospective data annotated by the radiologist. These results are quite good and indicate that either the SMO or the Bayesian classifier generated in this study can be well applied as a support tool to help physicians/radiologists to classify mass density in mammograms.

The values of $K$ and F-measure for this experiment are not so good as the ones obtained with the classifiers that predict outcome. The K value, once more, indicates that both NaiveBayes and SMO have a moderate level of agreement.

Performance of the best classifiers on unseen data Table 1 summarizes the results of predicting outcome on the 168 unseen cases as well as the results of filling up the attribute mass density in the test set. The first two lines of Table 1 refer to experiments to fill up values of the attribute mass density in the test set. The CCI indicates how well models $M_1$ and $M_2$, obtained respectively with experiments $E_3$ and $E_5$, performed on filling up those values, when compared with the actual values of retro_density available in the test set. The SMO classifier, which had a very good performance on the training set (CCI=81.3%), behaves even better when filling up values for retro_density, making mistakes in only 16% of the actual masses. The NaiveBayes classifier ($M_3$), obtained with experiment $E_3$, which had CCI=67.2% in the training set, performed very well in the task of filling up the missing values of density_num, correctly classifying 75.6% of the instances. A result that surpasses the performance obtained by the specialist, which is 70%.

For the tasks of predicting outcome, the classifiers also perform very well, with the worst predictions being produced by model $M_4$, which does not use any information about mass density. This result confirms once more the relevance of mass density on predicting outcome. In the absence of this information, the data could be filled up by $M_4$ or $M_5$, that, as mentioned, have a good performance on performing this job.

4 Conclusions and Future Work

In this work, we were provided with 348 cases of patients that went through mammography screening and biopsies. The objective of this work was twofold: i) find non trivial relations among attributes by applying machine learning techniques to these data, and; ii) learn models that could help medical doctors to quickly assess mammograms.

The conclusions are threefold: (1) automatic classification of a mammography, independent on information about mass density, can reach equal or better results than the classification performed by a physician; (2) mass density seems to be a good indicator of malignancy, as previous studies suggested; (3) machine learning classifiers can predict mass density with a quality as good as the specialist blind to biopsy, which is one of our main contributions. Our classifier can predict malignancy in the absence of the mass density attribute, since we can fill up this attribute using our mass density predictor.

As future work, we plan to extend this work to larger data sets, and apply other machine learning techniques based on statistical relational learning, since classifiers that fall in this category provide a good explanation of the predicted outcomes as well as can consider the relationship among mammograms of the same patient. We would also like to investigate how other attributes can affect malignancy or are related to the other attributes. Yet another stream would be to study why the parameter variation in the WEKA algorithms has a strong impact on the performance of the classifiers. Another important step forward would be to investigate with the physician, why some instances are consistently misclassified by all algorithms.

Acknowledgments

This work has been partially supported by the projects HURUS (PTDC/EIA-EIA/100897/2008) and DigScope (PTDC/EIA-CCO/100844/2008) and by the Fundação para a Ciência e Tecnologia (FCT/Portugal).

References


MammoClass Application The best models were integrated into MammClass. The application is freely available at http://cracs.fc.up.pt/mammoclass.
Abstract

Cardiovascular diseases have a high impact worldwide, especially in developing countries. Obesity in children has been associated with these pathologies which are only diagnosed later in their adulthood. However, only recently it has been identified a strong relation between hypertension in childhood and cardiovascular diseases in the adulthood. An early detection is therefore paramount being only constrained by the required effort to the medical doctor in analysing different set of tables in order to determine a positive pathological condition or not. A computer aided diagnosis (CAD) system can easily fill this gap. A preliminary study is here conducted on a dataset composed by a little more than 5 thousand patients. The goal is to objectively identify the most important characteristics that should be considered by an automated learner in a routine examination in order to predict hypertension in a patient.

1 Introduction

Cardiovascular diseases (CVDs) have a high death rate, especially in developing countries. It is well known that CVDs are associated to unhealthy lifestyles which govern most of populations in developed countries. Tobacco, alcohol, high blood pressure and obesity are the most common risk factors in these countries. Developing countries face other risk factors as underweight which contribute significantly to mortality by CVD [1]. It was already being identified that even though CVDs typically diagnosed in adults from developed countries are linked to their lifestyles, it is in the childhood that the first glimpses of CVDs appear through hypertension [3]. Traditionally, hypertension is identified through continuous screenings by measuring blood pressures and by filling short questionnaires. One way to determine if a patient has hypertension is by some rule of thumb. However, such rules are inadequate or imprecise especially in young children mostly due to the heterogeneous populations. Moreover, rules and procedures to assess hypertension for a given population are different and vary from different countries in the world. Therefore, panel experts gather periodically to define charts or guidelines. For this work, the adopted guideline is the one as referred in [2].

A computer aided diagnosis (CAD) system can assist the medical doctor to efficiently detect the presence or absent of hypertension in a patient. Moreover, it would also benefit the routine examinations by providing interpretable diagnosis so that a proper treatment could be prescribed to the patient. In fact, literature presents different comparative methodologies to predict hypertension. However, some are too complex or do not provide interpretable results [4], or do not take into count all available information [3].

In this work we study a data set consisted in more than 5 thousand children collected during a 7 year period. The objectives are three fold. We want to assess the key factors (age, height, weight, etc.) that linearly aid in determining if a given child has hypertension or not; Based on this reasoning, we want to objectively quantify if the models can accurately predict the patient condition; And, if the learning models can avoid misdiagnosis pathological patients as healthy. The structure of this work is as follows: Section 2 we present a detailed description of the data set used in this study and in Section 3 we describe our study followed by a discussion of the obtained results in Section 4. Conclusions are drawn in Section 5.

2 Data Set

The data set used in this study is encompassed by little more than 5 thousand patients collected at the Unidade Cardiológica Materno Fetal at Recife, Brazil during a 7 year period. 2173 are female and 3115 are male. Data dispersion in terms of age is as follows. It contains 1579 at preschool level (2–6 years), 1818 at school level (6–10 years), 1197 pre-pubescent (girls from 10–12 years and boys from 10–14 years), 455 pubescent (girls from 12–14 years and boys from 14–16 years) and 239 post-pubescent (girls from 14–18 years and boys from 16–20 years). Besides the normal information (such as weight, height, age and bmi–body mass index), we have also collected data regarding pulses and blood pressures (systolic and diastolic, sbp and dbp respectively) and cardiac frequency (cf). It was also possible to acquire the n/abn (normal vs. abnormal) information describing the relation between arterial pressure and age; b2 and murmur heart sounds are also present in this data set. Finally, the history of the present illness (hpi) and cardiac frequency (cf) were also recorded. To assess the pathological condition of a given patient, the arterial blood percentile pressure (app) was calculated according the information from [2]. app is coded by three levels of arterial blood percentile pressure: normal cases, pre-hypertension and hypertension. For this study we merged the pathological cases into a single one. Finally, from the aforementioned data set, 705 patients have hypertension. The remaining cases did not register any pathological condition. Table 1 summarizes the information regarding to this data set.

An important aspect is that gender is usually stated as one of the key factors in cardiovascular studies in children [1]. For better understanding of the study here conducted, we separated data presented in Table 1 by gender. After a careful analysis, we can check that the recorded features which compose this data set are strongly heterogeneous between genders. To assess this, we performed a χ2 test for homogeneity over all features independently. For instance, for the cardiac frequency, data was firstly ordered and divided into four quartiles for both genders. We then applied the χ2 test for homogeneity. For all features, the only ones indicated as homogeneous were the weight and age.

3 Feature Analysis through Logistic Regression

In this Section we will describe succinctly some concepts regarding to the logistic regression in this study, but first we will shortly present the

<table>
<thead>
<tr>
<th>Feature</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min.</td>
<td>Max.</td>
</tr>
<tr>
<td>weight (kg)</td>
<td>9.30</td>
<td>105.00</td>
</tr>
<tr>
<td>height (m)</td>
<td>0.70</td>
<td>1.92</td>
</tr>
<tr>
<td>bmi</td>
<td>4.90</td>
<td>41.50</td>
</tr>
<tr>
<td>age</td>
<td>2.06</td>
<td>17.98</td>
</tr>
<tr>
<td>sbp (mmHg)</td>
<td>70.00</td>
<td>160.00</td>
</tr>
<tr>
<td>dbp (mmHg)</td>
<td>40.00</td>
<td>120.00</td>
</tr>
<tr>
<td>cf</td>
<td>48.00</td>
<td>150.00</td>
</tr>
</tbody>
</table>

Table 1: Illustrates the variance values of this data set divided by gender. It is consisted, among others, by weight, height, body mass index (bmi), systolic and diastolic blood pressures (sbp and dbp, respectively) and cardiac frequency (cf). In each column it is presented the minimum, average and maximum values for each recorded feature.
nomenclature. Consider $x$ to be the given data comprising $N$ patient records, with input vector $x_1, \ldots, x_p, x \in \mathbb{R}^p$, and with the respective clinical condition values $y_1, \ldots, y_N$ where $y_n \in \{1, 2, 3\}$, i.e., $y_n$ is a normal, pre-hypertensive or hypertensive. For this study we merged the pathological cases into a single one.

### 3.1 Logistic Regression

Logistic regression is a statistical learning tool that permits to infer a linear relation between data and its outcome by modelling the posterior probabilities of the $K$ classes. Logistic regression usually takes the following form:

$$
\log \frac{P(C_i|x)}{P(C_j|x)} = \beta_0 + \beta^T x
$$

with $P(C_i|x) = \exp(\beta_0 + \beta^T x) / (1 + \exp(\beta_0 + \beta^T x))$ being therefore constructed by the monotone transformation function logit $p/(1 + p)$. Data analysis consist in one of the many resources of the logistic regression where it allows to understand the impact that each feature has explicitly on the outcome. Such characteristic will aid us in our experimental study which will further extended with the methods that we briefly describe next.

#### 3.2 Feature Analysis

In this Section we will use the learning method previously described to perform three main analyses. Using only the linear model trained with a sufficiently large sample, we will infer a linear relation among features and outcome. The predominant relations will indicate which features are linearly more relevant for the model. Afterwards, and based on the knowledge gathered we will quantify the performance gains.

One of the logistic regression characteristics is that it allow us to understand the relevance of each feature during the model construction. Therefore, we conducted an analysis of the coefficients of a learner to assess the most significant features. To do this, we applied a significance test by calculating its $Z$ score with 95% of confidence over 60% of the data set. By performing feature selection, it clearly benefits the prediction by attaining higher sensitivity rates.

5294 patients, logistic regression misses in performing the correct diagnosis in 317 patients. This result can be properly analysed by considering two possible errors that may be occurring. Knowing the importance on identifying normal (or benign) cases as truly normal (benign) cases or, in the same way, on identifying pathological (or malign) cases as truly pathological (malign) cases, we considered the sensitivity and specificity measure.

Values presented on Table 4 where calculated based on the performance analysis discussed on the previous section. It can be observable that logistic regression achieved higher sensitivity rates when using the pre-selected features. This has to do with some of the existing features that could erroneously bias our model.

### 5 Conclusion

It has been stated in the literature that an early detection of hypertension on children can avoid or diminish the probability to contract cardiovascular diseases in the adulthood. Standard routines involve a considerable amount of time to medical doctors. They encompass, for instance, patients’ record and historical data analysis to determine the health condition of a child. A CAD system can lessen this burden by providing an effective way to discriminate normal patients from hypertensive to allow a more intensive monitoring of risk groups. Another advantage that this system can provide is the ability to reason the pathological condition of a patient. In this work we performed a preliminary study on the detection of hypertension in children and which features are more relevant for the automatic learner. We have also assessed their performance and measure their effectiveness in avoiding false negatives.

### Acknowledgment

We would like to acknowledge to Fundação para a Ciência e a Tecnologia (FCT) through project PTDC/EIA-CCO/109982/2009.

### References


<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmi percentile</td>
<td>low weight, normal, overweight, obesity</td>
</tr>
<tr>
<td>pulses</td>
<td>normal, femoral diminished, diminished, high</td>
</tr>
<tr>
<td>n/abn</td>
<td>normal, abnormal</td>
</tr>
<tr>
<td>b2</td>
<td>normal, unique, fixed splitting, hyperphonic</td>
</tr>
<tr>
<td>murmurs</td>
<td>none, continuous, distolic, systolic</td>
</tr>
<tr>
<td>hpi</td>
<td>check-up, other, cardiological assessment, suspected cardiopathy,</td>
</tr>
</tbody>
</table>

Table 2: Illustrates the values for discrete variables in this data set. hpi is the history of the present illness.

<table>
<thead>
<tr>
<th>Training Sizes</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>90.7% ± 0.04</td>
<td>93.0% ± 0.02</td>
<td>91.3% ± 0.04</td>
</tr>
<tr>
<td>SF</td>
<td>94.6% ± 0.00</td>
<td>94.9% ± 0.00</td>
<td>94.9% ± 0.00</td>
</tr>
</tbody>
</table>

Table 3: Performance rates (and respective standard deviations) on 50 simulations with all (all) and selected features (SF).

<table>
<thead>
<tr>
<th></th>
<th>All Features</th>
<th>Selected Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.39 (0.31)</td>
<td>0.76 (0.02)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.99 (0.01)</td>
<td>0.98 (0.00)</td>
</tr>
</tbody>
</table>

Table 4: Shows the results for the sensitivity and specificity rates when using all features and model pre-selected features when training with 60% of the data set. By performing feature selection, it clearly benefits the prediction by attaining higher sensitivity rates.
Detection of Carotid Plaque Symptoms using Ultrasound Imaging

David Afonso
dafonso@isr.ist.utl.pt
J. Miguel Sanches
www.isr.ist.utl.pt/~jms

This work was supported by the FCT project (PEst-OE/EE/UI0009/2011)

Abstract

Carotid plaques are one of the commonest causes of neurological symptoms due to embolization of plaque components or flow reduction. The classification of plaques vulnerability is then a relevant clinical issue, and a technical challenge. Recently, several atherosclerotic plaque characterization methods were proposed based on plaque morphology assessed through 2D ultrasound. One of these methods, proposed by Seabra et al. presents a measure with clinical significance, known as enhanced activity index (EAI), that the clinician then uses to classify the plaque.

The present paper builds upon that work and by using machine learning, proposes an ensemble classifier that shows promising results globally outperforming both gold medical standard degree of stenosis and EAI. Results are obtained on a real clinical database of 146 plaques of which 44 of them are symptomatic.

Future work will investigate the predictive performance of the proposed classifier, i.e., how well does the classifier identify stable lesions at high risk of becoming symptomatic.

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 easily cause myocardial infarction, stroke and lower limb ischemia. Correct characterization of the carotid disease is then 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.

The degree of stenosis (arterial lumen narrowing) is up to now considered the most important features for determining the plaque vulnerability. This metric, together with other patient information such as age, health and clinical history are features clinicians usually use to subjectively decide upon endarterectomy. Even though, several numerous studies [1, 2, 3, 4] report that plaque morphology is also an important ultrasound marker that positively correlates with symptoms.

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, it is non-invasive, does not involve ionizing radiation, it is cheap and is very common in clinical facilities.

There are several recently proposed methods [1, 2, 4, 5], that make use of plaque morphology to characterize carotid plaques. In general it is argued that an optimal method for identifying vulnerable lesions should include morphological and textural features, extracted from pixel intensity information, and clinical information regarding plaque structure and appearance (e.g. stenosis, evidence of surface disruption and presence of echogenic cap) given by experienced physicians. The combination of this information is expected to produce a more comprehensive description of the profile of a vulnerable plaque.

J. Seabra method [5] makes use of very interesting texture features, by not discarding the so called “image noise”, but rather considering this as an important information source of tissue texture, translated into US speckle. It then proposes a measure with clinical significance, known as enhanced activity index (EAI), that the clinical then uses for classification.

Although this method might be comfortable for the clinicians, because it provides them with a score that they are used to having (such as the degree of stenosis, age, etc.) to them make a decision upon, it restricts the classifier performance and hence it’s performance.

In this work we propose a full machine learned classifier that makes it possible to identify a binary detection of plaque vulnerability, and that outperforms both the degree of stenosis (DS) gold-standard and the EAI score method.

2 Problem Formulation and Data

We wish to develop a classification method to decide if a carotid lesion (plaque) is associated to neurological events (symptoms). A plaque was considered symptomatic when amaurosis fugax (AF) or focal (transitory, reversible or established) neurological symptoms in the carotid territory were observed in the previous 6 months. We then have a binary detection problem, where we define the positive class (P) as symptomatic and the negative class (N) as asymptomatic. 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 \( P = 44(30\%) \).

3 Methods

We employed a three-step method sketched in Fig. 1. The first step (i) consists of feature extraction from the US images. The second (ii) step consists of training and evaluating several classifiers in the full feature space. In the third (iii) stage a final classifier ensemble is build and evaluated on a reduced feature space.

![Figure 1: Scheme of the phases (i, ii, iii) implemented to build and evaluate the proposed classifier.](image-url)
3.1 Feature extraction (i)

The final dataset is comprised of two subsets of features: A) the 4 features obtained by the clinician (existence of fibrous cap, surface disruption, plaque texture homogeneity and DS) which are mainly binary and B) 110 features computationally extracted from the images, via a series of post-processing steps (histogram features, Rayleigh mixture models features, Rayleigh parameters, textural features, morphological features, etc) [5] consisted of real numbers.

3.2 Classifier evaluation and selection (ii)

We use the PRTools Matlab toolbox to train and test 18 different classifiers on the separate subsets A and B, using a bootstrap approach. Although we do not present comparison results, this dataset separation lead to better results, presumably because of their different nature. Almost unbiased metrics of Probability of error, Sensitivity, Specificity and Precision are computed using the .638 bootstrap estimate:

\[
\text{metric}_{632} = 0.382 \times \text{metric}_{	ext{Apparent}} + 0.632 \times \text{metric}_{	ext{Jack-knifing}}
\]

where the apparent (optimistic) estimate has equal training and test sets, and Jack-knifing is a pessimistic estimator where training and sets are disjoint. All sets are randomly selected, and metrics averaged over 100 bootstrap repetitions.

As expected, several classifier have similar performance on dataset A, due to its low dimensionality, including the Support Vector Classifier. On the other hand on dataset B the classifiers performances diverge. We then construct a stacked ensemble of the top 5 classifiers on B with the lowest difference between sensitivity and specificity and below average probability of error (see Table 1). Besides its intrinsic logic, the main reason for using this criteria is that the method in [5] also proposes the same "sensitivity = specificity" to set the final threshold over the EAI score for classification purposes, and we wish to have a fair comparison as much as possible.

The classifiers combination rule follows the class which yields the highest value of the product of the classifiers posterior probabilities.

3.3 Feature Selection and classifier ensemble (iii)

To avoid classifier overfitting, feature selection reduced the size of subset B from 110 to 11 features (subset C). This dimension was optimally determined by the forward-search procedure as the feature space that lead to the lowest probability of error, taking the 1-Nearest Neighbor criteria. The final classifier is a parallel ensemble of the stacked classifiers (see Table 1) on dataset C and the Support Vector Classifier on dataset A. The combination follows the class with the highest vote of the base classifiers by using:

\[ D(k, j) = (v + 1)/(n + c) \]

where \( v \) is the number of votes object \( k \) receives for class \( j \), \( n \) is the total number of classifiers, and \( c \) the number of classes (two). The training and evaluating procedures are the same as described in section 3.2.

4 Experimental Results

The performance results of the individual classifiers on both subsets A and B, discussed in section 3.2, are display on Table 1.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Set</th>
<th>Prod. Error</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic Discriminant</td>
<td>B</td>
<td>0.53 ± 0.31</td>
<td>65.97 ± 1.11</td>
<td>70.97 ± 0.47</td>
<td>55.47 ± 0.37</td>
</tr>
<tr>
<td>Fisher</td>
<td>B</td>
<td>30.91 ± 0.27</td>
<td>65.51 ± 1.27</td>
<td>70.67 ± 0.43</td>
<td>55.39 ± 0.36</td>
</tr>
<tr>
<td>Logistic</td>
<td>B</td>
<td>29.26 ± 0.28</td>
<td>66.90 ± 1.11</td>
<td>72.44 ± 0.42</td>
<td>57.37 ± 0.38</td>
</tr>
<tr>
<td>Linear</td>
<td>B</td>
<td>26.99 ± 0.20</td>
<td>60.23 ± 1.14</td>
<td>88.96 ± 1.30</td>
<td>53.55 ± 0.91</td>
</tr>
<tr>
<td>Support Vector</td>
<td>B</td>
<td>16.04 ± 0.19</td>
<td>73.36 ± 1.11</td>
<td>88.11 ± 0.25</td>
<td>72.96 ± 0.75</td>
</tr>
</tbody>
</table>

Table 1: Individual classifier performances on the subsets A and B. All metrics are .638 corrected for a more unbiased estimate. Only one classifier for A is shown because several classifiers displayed very similar behaviour.

The performance of the classifiers on dataset B is clearly not as good as on A, as expected. The stacked classifier ensemble help then to dilute their errors and obtain a better predictive performance, as we can see in Fig. 2.

These final results (Fig. 2) show a relative reduction of error probability of 61% over DS and 56% over EAI. Precision also displays better performance with an increase of 41% and 36% over DS and EAI, respectively. Although Specificity also shows a performance with a 21% relative increase over both DS and EAI, there is a trade-off in Sensitivity where there is a decrease of 8% and 7% respectively.

5 Conclusions and Future work

We present a work-in-progress that shows promising results in the detection of asymptomatic v.s. symptomatic (vulnerable) carotid plaques, using a machine learned classifier ensemble over a set of features extracted from ultrasound images by both the clinician (A) and computer processing (B). We approach the problem by building two expert classifiers on both (A) and (B) datasets, and ensemble them in a parallel classifier. To build these experts we search for good performance individual classifiers out of 18 on these subsets, and ensemble them in a stacked approach for dataset (B).

After feature selection, to avoid overfitting, final results show a significant increase in performance on error probability, specificity and precision when compared to the degree-of-stenosis (which is the medical standard) and the EAI method proposed in [5]. But there is still a decrease in Sensibility which does not allow us a clear statement of over performance. Future work will be focusing on this sensitivity issue.

This detection procedure allows a characterization of the plaque vulnerability. Future work will provide an estimate of the predictive power (indentify asymptomatic plaques that develop symptoms down the line) of this characterization on a longitudinal study.

References


Abstract

Carotid atherosclerosis diagnosis is made based on patient history and ultrasonography imaging. Understanding how the prevalence of this disease in the target population may affect the decision threshold in ROC curve analyses and provide an enhanced and more objective decision process for carotid atherosclerosis diagnosis and treatment is addressed here. Previous work has shown that strategies based on scores, like the Youden index, well known and vastly used in medical diagnostic tools, can be adapted to a prevalence dependent analysis, as well as Descending Diagonal Interception Criterion (DDIC). It is introduced to the referred methods a weight factor $\lambda$, that allows the possibility of weighting differently the TPR and the FPR. In general medical applications, the cost associated with True Positive Rate (TPR) is different from False Positive Rate (FPR), hence weighting them differently is a useful feature. It is graphically shown the influence of this factor in medical decision making when selecting patients for endarterectomy, as different cut-off values arise.

1 Introduction

It is clear that medical decision sustained by Computer-Aided Design (CAD) tools and other support tools for the decision making process, will allow better reproducibility and objective results. The diagnostic performance of a test, or the accuracy of a test to discriminate diseased cases from normal cases is evaluated using Receiver Operating Characteristics (ROC) curve analysis [6]. They are also used to compare the diagnostic performance of two or more classifiers. Hence, a worthwhile analysis about sensitivity and specificity (for each decision point) is a key aspect of our methodology. The principles of varied methodologies that are found in the literature addressing ROC curve analysis do evaluate a given classifier for its performance. Though, none of these methodologies provided a method answering to a priori knowledge of population behavior or incidence. It is also important to consider a more flexible evaluation of patients, given the physiologist experience, time line, economical and logistical aspects.

2 Problem Formulation

Our goal is to correctly classify a patient into specific populations and consequently, choose the appropriate decision threshold value for the decision making process. This decision point will, from now on, be referred as cut-off. For every possible cut-off point or criterion value you select to discriminate between the two populations, there will be some cases with the disease correctly classified as positive (TPR = True Positive fraction), but some cases where the disease will be classified as negative (FNR = False Negative fraction). On the other hand, some cases without the disease will be correctly classified as negative (TNR = True Negative fraction), but some cases without the disease will be classified as positive (FPR = False Positive fraction) [3]. The referred classifications are usually presented in a contingency table, also called confusion matrix [1].

The Sensitivity of a diagnostic test, also called True Positive Rate (TPR), is defined as $TPR = \frac{TP}{TP+FN}$. It can be understood as the probability of classifying a subject as positive, when the disease is present. Specificity [5] is defined as $Specificity = 1 - FPR$ where $FPR = \frac{FP}{TN+FP}$. It is the probability that a test result will be negative when the disease is not present.

In the specific case of endarterectomy, a binary classification problem, two classes are considered: $\omega_0$ and $\omega_1$ containing the subjects that will present symptoms in the future and that will continue asymptomatic, respectively. The main goal concerning endarterectomy is to decide what class a given subject belongs to. Based on the respective Enhanced Activity Index (EAI) score patients are classified as symptomatic or asymptomatic. Hence, it is expected with real data that the distributions overlap, as one rarely observes a perfect separation between the two groups.

A close analysis of cut-off criteria is necessary in order to correctly decide which criterion is most appropriated for different situations. The Descending Diagonal Interception Criterion (DDIC) and Youden’s criterion, already addressed in [2] are two examples of how one can select the most appropriate cut-off. They will now be analysed in a graphical receiver operating characteristic approach. For the present work, it is necessary to understand cut-off selection by the need of correctly classifying patients as symptomatic or asymptomatic, but also, to adjust that classification with economical and logistical aspects. We introduce a possibility that considers different weighing in Sensitivity(TPR) and Specificity (1-FPR) values, providing adjusted cut-off values.

3 ROC Geometrical Approach

ROC curves are used in medicine as a way to analyse the performance of diagnostic tests. This allows the determination of the cut-off value for distinguishing, for example, between positive and negative test results. Diagnostic testing is almost always a tradeoff between sensitivity and specificity and ROC curves provide a graphical representation of this tradeoff [4].

Since both Sensitivity, $Sens(c)$, and Specificity, $Spec(c)$, depend on the cut-off parameter, the ROC curve can be defined as the following parameter curve, $s(c): R \rightarrow R^2$, where $s(c) = (1 - Spec(c), Sens(c))$. We plot a point representing these rates on a two dimensional graph. The graphical representation of these sensitivity/specificity pairs, when the densities are fixed but the cut-off $c$ is changed, represent the ROC curve.

To better understand the dynamic of ROC plots, Figure 1 presents four points of analysis. Together, these points represent extreme situations of ROC analysis, gathering also the best and worst case scenarios. Point a) is the best possible result. It represents a test with perfect discrimination with a ROC curve that passes trough the upper left corner (Sensitivity=1 and Specificity=1). A curve that passes through point c) misses all classifications (Sensitivity=0 and Specificity=0). It represents a classifier that indicates a healthy patient as diseased, and a diseased patient as healthy. Point b) represents a classification of all individuals as symptomatic, regardless of score (Sensibility=1 and Specificity=0). Finally, point d) represents a classification of all patients as asymptomatic, regardless of score (Sensibility=0 and Specificity=1). Therefore, the closer
the ROC curve is to the upper left corner, the higher is the overall accuracy of the test [6]. The problem under analysis is about Sensitivity/Specificity values between 0 and 1, that results in the representation of a curve which demands further evaluation criteria for the selection of the appropriate cut-off value. Criteria for the computation of an optimal cut-off value are described next.

**Descending Diagonal Intersection Criterion (DDIC)**

From the ROC perspective, this criteria sets the optimal cut-off in the interception point between the ROC curve and the descending diagonal. This criterion may be formulated as

$$TPR + FPR = 1 \quad (1)$$

The interception of the ROC curve with the line at 90 degrees to the no-discrimination line is considered as an heuristic method to investigate the cut-off providing the best discriminative power of the test (or predictive method), maximizing the TPR and minimizing the FPR. In general medical applications, the cost associated with TPR is different from FPR, hence weighting them differently is a useful feature. Accordingly, let us consider the cost variable $\lambda$ in our criterion, and expand equation 1, such that

$$\lambda \times TPR + (1 - \lambda) \times FPR = \lambda \quad (2)$$

This means that the tangent to the ROC curve at the optimal cut-off, under the Youden criterion, has an unitary slope.

Once again, in order to consider adjustments to cut-off selection, $\lambda$ is considered in the criterion as a weighting factor. Eq. 3 can be written as:

$$\frac{d}{dc}(TPR(c)) = \frac{1}{\lambda} \quad (4)$$

Figure 3 is the graphical representation of the result presented by Eq. 4. For different relative values of TPR and FPR, Youden’s consideration for cut-off selection will behave according to different slopes for the tangent to the cut-off point on the ROC curve.

4 Discussion

The optimal cut-off under both criteria is different. In geometrical terms the DDIC optimal cut-off, $c^*$, is obtained by intersecting the ROC curve with the descendant diagonal while in the Youden criterion, the optimal cut-off occurs when the tangent to the ROC curve is unitary. When considering different weights for $\lambda$, DDIC will select the cut-off value where the ROC curve intersects the descending line with negative slope indicated by $y = 1 - \frac{1}{\lambda}x$. In the case of Youden’s index, the line tangent to the optimal cut-off point in the ROC curve is given by $\frac{dy}{dx} = \frac{1}{\lambda}$. From the data presented by J. Seabra et al. [5] and the use of the Enhanced Activity Index classifier, we show in Table 1, the influence in the cut-off selection from different $\lambda$ values, resulting also in different Sensitivity and Specificity:

<table>
<thead>
<tr>
<th>Lambda</th>
<th>DDIC</th>
<th>Youden</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4</td>
<td>0.61358</td>
<td>0.75358</td>
</tr>
<tr>
<td>3/8</td>
<td>0.69231</td>
<td>0.76923</td>
</tr>
<tr>
<td>1/2</td>
<td>0.76923</td>
<td>0.84615</td>
</tr>
<tr>
<td>5/7</td>
<td>0.84615</td>
<td>0.90888</td>
</tr>
<tr>
<td>10/11</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Values of lambda and respective sensitivity, specificity and cut-off values for DDIC and Youden Index.

5 Conclusions

In population selection, the inclusion of different weight values $\lambda$ can be interpreted as a wider, or narrower, selection of the symptomatic population, resulting in bigger or smaller values of TP fraction. Like this, $\lambda$ represents an adjustment that can be made to the initial criterion. A narrower and less demanding selection of TP is given by smaller values of $\lambda$. Bigger values of $\lambda$ will represent a wider selection of the symptomatic population (resulting in more TP). When applying these thoughts to the endarterectomy situation, one would want to have bigger values of $\lambda$ for situations where less resources (human, economical or even logistical) are available. The opposite is also true, allowing with lower values of $\lambda$ more patients to be submitted to surgery. Like this, further work is needed in order to understand how this parameter would result as a cost function for endarterectomy decision making.

References


Facial Dynamics for Identity and Expression Recognition

Carolina Raposo
Jorge Batista
http://www2.1sr.uc.pt/~batista/

Abstract

Psychological studies indicate that facial dynamics is a biometric, i.e., it can be used for identity recognition. Based on this information, the present work attempts to demonstrate that facial motion alone is sufficient for performing person identification. Besides identity recognition, expression recognition is also performed. The work developed includes the usage of a facial descriptor which makes use of dense texture information, Volume Local Binary Patterns (VLBP). Since it is desirable to analyse only the dynamic components, image subtraction is used for removing shape and texture information. Moreover, tensors are used for performing data analysis. The novelty in this work includes the development of a procedure which combines existing techniques in a new way. In general, all the experiments yielded good results, demonstrating that facial dynamics is a proper biometric. Confirmation was obtained since it has been shown that it is possible to identify an individual whose appearance is significantly different from the original.

1 Introduction

Identity and expression recognition has been a branch of computer vision with growing importance in the past decades. The term “biometrics” refers to the measurable biological characteristics which are used to quantify the physical features of an individual for use as a means of identification. A relatively new area of study is the dynamics of facial expression. The term “dynamics”, in this context, can be defined as the changes in facial motion over sequential time. One of the advantages over static analysis is that facial dynamics is less affected by physical changes such as ageing, gaining weight, wearing glasses, growing a beard, etc. Studies show that body and facial movements support person identification. There is considerable evidence that dynamic information is not redundant and may be beneficial for various aspects of face processing, including age, gender, and identity recognition ([3], [5]). Based on this knowledge, the main objective of this work is to corroborate the following hypothesis: facial expressions can be used as an effective biometric for person identification. The growing importance of facial dynamics as a field of research, in both identity and expression recognition, is due to its many applications and to the fact that it constitutes a relevant biometric.

The overall methodology of the procedure presented in this work includes the usage of a dense facial descriptor (VLBP) of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise), a technique for reducing shape and texture information (image subtraction) and a data analysis method (decomposition of tensors).

2 Volume Local Binary Patterns

Volume Local Binary Patterns (VLBP) [2] are an extension of the classic Local Binary Patterns (LBP) to perform dynamic texture analysis and represent temporal sequences as vectors. In general terms, the LBP operator forms labels for the image pixels and creates a histogram using these labels as a texture descriptor. Firstly, the neighbourhood of a pixel, composed of $P$ equally spaced pixels on a circle of radius $R$, is thresholded with the value of that pixel and a binary number is assigned to it. Afterwards, for every pixel in the neighbourhood, a weight is defined. By summing the multiplied weights and binary values, an LBP code is obtained for each pixel. Finally, the occurrences of the LBP codes in the image are collected into a histogram which is the texture descriptor of the image. The LBP operator only deals with spatial information. To incorporate temporal information, VLBP can be used. The idea is that instead of considering a 2D neighbourhood, the face sequence is seen as a rectangular volume and the neighbourhood of each pixel is defined in three dimensional space. In this work, each face in each frame is detected, using a software for locating faces, and aligned to a reference frame. Alignment of the faces is needed because VLBP computes the codes considering the neighbourhood of each pixel, in sequential frames, and thus translation, rotation and scale components must be removed.

3 Tensors

Tensors [1] are multidimensional arrays of data and therefore can be used to analyse multivariate data. A vector is considered as a first-order tensor and a matrix as a second-order tensor. In this work tensors are used to describe dynamic sequences of different people performing different expressions. Extending the concept of matrix SVD, Higher Order SVD (HOSVD) decomposes an $N$-order tensor $\mathcal{D}$ into $N$ orthogonal spaces $U_1 U_2 \ldots U_N$ and expresses $\mathcal{D}$ as the mode-$n$ product: $\mathcal{D} = Z \times_1 U_1 \times_2 U_2 \ldots \times_N U_N$, where $Z$ is the core tensor. In this work, HOSVD is used as in [1]. It is beneficial since it allows a multi-factorized space to be decomposed into its constituent modes. The different modes can then be analysed separately and important information about the data can be extracted. A database with a significant amount of individuals may comprise a great variation of textures (due to different ethnicity), lighting conditions, poses, genders, facial expressions, etc. It is often desirable to analyse only a certain characteristic (e.g. the gender) in the presence of variations of other characteristics. Thus, a technique for separating the different characteristics is of great use, so that only the target one is considered. Tensor analysis can be used for solving this problem, since it provides a separation of the data in subspaces representative of each characteristic. In this work, identity and facial expressions are the two characteristics subject to analysis.

4 Frame Subtraction

Since the main objective of this work is to demonstrate that facial dynamics is a biometric, experiments are performed not only using the original data sets but also modified data sets which attempt to eliminate or reduce the component related to facial shape and texture by subtracting the neutral face of each individual from every frame in the database. The result for a sequence of frames is shown in figure 1, where it can be seen that only the areas which exhibit greater dynamism are depicted.

Figure 1: Frames from the original (top) and the subtracted (bottom) databases.

5 Identity and Expression Recognition

Each face is represented by a grey scale image of $40 \times 40$ pixels. A software for locating faces which removes translation and scale components is used. Rotation components are removed using the positions of the eyes. Previous work [1] has shown that different streams of a certain facial expression, which are short sequences representing part of the facial expression, have similarities in the sense that when represented in the same referential, they cluster together. These streams are constructed by sampling each sequence using a windowing technique such that multiple samples of each sequence are ascertained. VLBP is the used texture descriptor so that each stream is represented by an histogram of grey levels. Each stream is created by concatenating histograms obtained after
applying VLBP to $M \times N$ volumes, constructed by dividing each frame in $M \times N$ non-overlapping blocks. Since the histograms encode texture and temporal information, a 3D-tensor is created, so that identity, facial expressions and VLBP code occurrences are represented in orthogonal subspaces. More specifically, for a 3D tensor $D \in \mathbb{R}^{I \times J \times K}$, $I$ is the number of people in the database, $J$ is the number of facial expressions (in this case it is 6) multiplied by the number of repetitions of each expression and $K$ is the dimension of each facial sequence. Using HOSVD, $D$ is expressed as the mode-$n$ product of a core tensor, $Z$, and 3 orthogonal subspace matrices, $U_{\text{people}}, U_{\text{expressions}}$ and $U_{\text{timeVec}}$.

For each subspace matrix, each row vector represents one element of the considered characteristic and the row vectors span the space of that characteristic from the database across the remaining ones. Each test vector is projected to people or expression subspaces, and identity and expression classification are performed by computing distances to the people or expression subspaces, and identity and temporal information, a 3D-tensor is created, so that identity, facial expressions and VLBP code occurrences are represented in orthogonal subspaces. More specifically, for a 3D tensor $D \in \mathbb{R}^{I \times J \times K}$, $I$ is the number of people in the database, $J$ is the number of facial expressions (in this case it is 6) multiplied by the number of repetitions of each expression and $K$ is the dimension of each facial sequence. Using HOSVD, $D$ is expressed as the mode-$n$ product of a core tensor, $Z$, and 3 orthogonal subspace matrices, $U_{\text{people}}, U_{\text{expressions}}$ and $U_{\text{timeVec}}$.

For each subspace matrix, each row vector represents one element of the considered characteristic and the row vectors span the space of that characteristic from the database across the remaining ones. Each test vector is projected to people or expression subspaces, and identity and expression classification are performed by computing distances to the people or expression subspaces, and identity and temporal information, a 3D-tensor is created, so that identity, facial expressions and VLBP code occurrences are represented in orthogonal subspaces. More specifically, for a 3D tensor $D \in \mathbb{R}^{I \times J \times K}$, $I$ is the number of people in the database, $J$ is the number of facial expressions (in this case it is 6) multiplied by the number of repetitions of each expression and $K$ is the dimension of each facial sequence. Using HOSVD, $D$ is expressed as the mode-$n$ product of a core tensor, $Z$, and 3 orthogonal subspace matrices, $U_{\text{people}}, U_{\text{expressions}}$ and $U_{\text{timeVec}}$.

For each subspace matrix, each row vector represents one element of the considered characteristic and the row vectors span the space of that characteristic from the database across the remaining ones. Each test vector is projected to people or expression subspaces, and identity and expression classification are performed by computing distances to the people or expression subspaces, and identity and temporal information, a 3D-tensor is created, so that identity, facial expressions and VLBP code occurrences are represented in orthogonal subspaces. More specifically, for a 3D tensor $D \in \mathbb{R}^{I \times J \times K}$, $I$ is the number of people in the database, $J$ is the number of facial expressions (in this case it is 6) multiplied by the number of repetitions of each expression and $K$ is the dimension of each facial sequence. Using HOSVD, $D$ is expressed as the mode-$n$ product of a core tensor, $Z$, and 3 orthogonal subspace matrices, $U_{\text{people}}, U_{\text{expressions}}$ and $U_{\text{timeVec}}$.

5.1 Experimental Results

Databases 1, 2 and 4 comprise 4, 29 and 1 people, respectively. Database 4 includes facial expressions with appearance changes.

5.1.1 Database 1

Using Database 1 to evaluate the effect of the number of blocks used, results are shown in Table 1. In theory, using more blocks leads to a more detailed description of the faces since the grey level histograms are created using smaller parts of the face. In practice, this can be confirmed by analysing the results which show that higher identity and expression recognition rates are obtained as the number of blocks increases. Comparing to the results obtained in [4], where identity and expression recognition are performed using Database 1, the improvement in both recognition rates is of 4%, approximately.

<table>
<thead>
<tr>
<th>No. of Blocks</th>
<th>Subtracted DB</th>
<th>Original DB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identity Recog</td>
<td>Expression Recog</td>
</tr>
<tr>
<td>1x1</td>
<td>76.04%</td>
<td>64.58%</td>
</tr>
<tr>
<td>3x2</td>
<td>91.67%</td>
<td>82.29%</td>
</tr>
<tr>
<td>5x4</td>
<td>98.96%</td>
<td>89.58%</td>
</tr>
</tbody>
</table>

Table 1: Results obtained with stream length 40 and step size 5 frames.

5.1.2 Database 2

Table 2 shows the results for varying stream lengths and step sizes, using Database 2. Smaller steps, which correspond to higher percentages of common frames between training and testing sequences (overlapping), lead to higher identity and expression recognition rates. In this case, even with very small or even non-existing overlapping, very high expression recognition results are obtained. Considering only expression recognition, experiments with Database 2 show that the difference between rates obtained with the subtracted and the original data sets is higher than when using Database 1. This is due to the fact that larger intersubject differences are present, resulting in worse expression recognition results when using the original data set. This can be confirmed by observing figure 2, which depicts the differences between expression recognition rates obtained with the subtracted and the original data sets, as the number of individuals increases. Results clearly show that, as the number of people in the database increases, the difference becomes larger.

<table>
<thead>
<tr>
<th>Length</th>
<th>25</th>
<th>35</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step</td>
<td>22 (11.2%)</td>
<td>20 (14.3%)</td>
<td>15 (6.7%)</td>
</tr>
<tr>
<td>% Identity Recog.</td>
<td>100 / 100</td>
<td>100 / 100</td>
<td>100 / 100</td>
</tr>
<tr>
<td>% Expression Recog.</td>
<td>91.38 / 97.13</td>
<td>94.33 / 99.43</td>
<td>97.91 / 99.43</td>
</tr>
</tbody>
</table>

Table 2: Results obtained using both the original (left values) and the subtracted (right values) databases. 3 x 2 blocks were used. The values between parenthesis indicate the maximum percentage of common frames between the query sequence and the training sequences.

5.1.3 Databases 1 and 4

In order to test the efficacy of this procedure in the presence of significant appearance changes, a new database was created, Database 4, in which an individual performs the six basic expressions, three times each, without and with appearance changes (using paint and foam). Experiments are performed by merging databases 1 and 4, so that 5 individuals are considered. The altered sequences of Database 4 are used for testing. Since it is desirable to assess the method as a biometric, only identity recognition results are shown.

![Figure 2: Difference between expression recognition rates obtained with the subtracted and the original data sets, for increasing number of people.](image)

Average identity recognition results obtained after varying stream lengths and step sizes are shown in table 3. It can be observed that even with the original data set, some streams of the subject’s painted face were correctly classified, which indicates that the subject is not completely recognisable. However, much higher rates were obtained when using the subtracted data set because the identification was performed by using mostly dynamic information. In this case, very good recognition results were achieved, proving that facial dynamics is definitely a useful biometric. For the sequences filmed with the face covered with foam, with the original data set, the significant change in appearance leads to incorrect classification for all the query streams. Reducing shape and texture components, leads to a great increase in the recognition rates. It has been demonstrated that facial dynamics is an efficient biometric since it is clearly possible to perform identity recognition even when the individual is “masked”.

6 Conclusion

This work demonstrates that facial dynamics, in the form of facial expressions, constitutes a biometric, i.e., can be utilised for performing identity recognition. It has been observed that, in the presence of a significant number of individuals, using dynamic information alone, by minimising or even removing shape and texture information, yields better recognition results. This indicates that the interpersonal differences, more evident in shape and texture cues, constitute disadvantageous information when encoding the different facial expressions. The relevance of the good results obtained with the new database is that they evince the fact that facial dynamics is indeed a proper biometric, because the individual cannot be identified using its texture and shape information alone.

References

Detecting Cardiac Pathologies from Annotated Auscultations

Pedro Ferreira\(^1\) pedroferreira@dcc.fc.up.pt
Inês Dutra\(^2\) and Miguel Coimbra\(^2\) ines@dcc.fc.up.pt, mcoimbra@dcc.fc.up.pt
Felipe Mourato\(^3\) and Sandra Mattos\(^3\) felipe.a.mourato@gmail.com, ssmatos@cardiol.br
Daniel Pereira\(^4\) and Ricardo Cruz-Correia\(^4\) danielclaudiopereira@gmail.com, rcorreia@med.up.pt

\(^1\) CRACS-INESC TEC, Porto, Portugal
\(^2\) CRACS-INESC TEC and IT-Porto
\(^3\) Real Hospital Português, Recife, Pernambuco, Brazil
\(^4\) CINTESIS, Porto, Portugal

Abstract
The DigiScope project aims at developing a digitally enhanced stethoscope capable of using state of the art technology in order to help physicians in their daily medical routine. In this work, we present a preliminary analysis and study of the first auscultations performed on children of a Brazilian hospital. Results indicate that classifiers can be obtained that distinguish reasonably well patients with cardiac pathologies from those that do not have pathologies.

1 Introduction
Digital stethoscopes are medical devices that can collect, store and sometimes transmit acoustic auscultation signals in a digital format. These can then be replayed, sent to a colleague for a second opinion, studied in detail after an auscultation, used for training or even can be used as a cheap powerful tool for screening cardiac pathologies.

DigiScope\([1,2]\) is one of the enhanced stethoscopes that aims at using state of the art technology in order to help physicians in their daily medical routine. DigiScope aims to be a prototype of a digitally enhanced stethoscope, capable of automatically extracting clinical features from the collected data, as well as providing a clinical second opinion on specific heart pathologies.

In this work, we used data collected by the DigiScope Data Collector, which is being used in two hospitals, one in Portugal and another one in Brazil\([1]\). The data collected in Portugal is from adults while the data collected from Brazil is from children. In this particular paper, we concentrate on analyzing the Brazilian children data. We give some statistics about the data being collected in Brazil and report preliminary results on the correlation among the data attributes that may lead to valuable recommendations to the doctors and their patients.

This work was approved by the Ethical Review Board of the Real Hospital Português.

It has two goals:
1. automatically learn classifiers that distinguish between normal patients and those with any cardiac pathology. Our classifiers rely only on the cardiologist provided annotation and not on the raw sound data itself.
2. automatically extract new and relevant knowledge from the dataset.

Very few works in the literature report on prediction of heart diseases using machine learning techniques. A recent work\([3]\) had as objective to model the detection of heart failure more than 6 months before the actual date of clinical diagnosis using machine learning techniques to Electronic Health Record (EHR) data. They compared the performances of logistic regression, SVM and Boosting along with various variable selection methods in heart failure prediction. They reported a value of 0.77 for the Area Under the ROC (AUROC) for the best classifier. With our dataset and performing an exhaustive search for the best classifier, we obtained an accuracy of 90.5% and AUROC of 0.83 on unseen cases.

2 Materials and Methods

2.1 Data
The DigiScope application has been used for four months at two hospitals. During this period, June to September 2011, around 200 patients (children) were auscultated at the Real Hospital Português, in Pernambuco, Brazil. Each auscultation was recorded, and a XML file containing the physician annotated data was associated with each sound file. In this work, we focus only on the annotated nominal data available in the XML files. For each patient, we have 40 attributes, but we only use the variables that were annotated for the majority of the annotations (i.e., those that have few missing values). These attributes are: Age, Height (cm), Weight (kg), SystolicSystemicPressure (mmHg), DiastolicSystemicPressure (mmHg), Sex, AuscultationPosition, SystemicPressureMethod, Murmur, S2Status, IFAbnormal, PulmonaryComponent, CardiacPathology, CardiacPathologyType, S1Status, PressurePosition, S3Exist, S4Exist, StatusForm.

The average age for the children in the study is 7.3 years.

We worked with the 169 cases that had almost complete annotations and were not missing the class label CardiacPathology. 40 cases were labeled as having a CardiacPathology while the remaining 129 were annotated as being normal. Besides being auscultated, every patient also had an ecocardiogram. The final label for CardiacPathology was determined during auscultation and possibly modified after the ecocardiogram.

2.2 Methodology
Besides the original attributes, we derived the following additional ones: body mass index (BMI, calculated as the person's weight in kilograms divided by the square of the height in meters), and categorized versions of SystolicSystemicPressure (mmHg) and DiastolicSystemicPressure (mmHg), according to the children's sex, age and respective height percentiles.

For training, we omitted the attributes related to weight, height and age, since these are captured by derived attributes generated, such as body mass index (BMI, calculated as the person's weight in kilograms divided by the square of the height in meters). We used categorized versions of the attributes in order to ensure that the discretized versions are medically relevant. No weighing was used to compensate the class size imbalance.

Our main goals with this work are: (1) to distinguish between abnormality (patients that have a cardiopathy) and normality (patients that do not have any cardiopathy), and (2) explore the data for new relevant knowledge. For the first goal we focused on learning a classifier to predict CardiacPathology using the other attributes. According to the specialists, the attribute Murmur is considered to be very important to predict a cardiac pathology. In order to study the influence of the attribute Murmur, we also performed experiments removing the attribute Murmur from our dataset. These experiments were performed using the WEKA tool\([4]\). We compared several classifiers with a variety of parameters. For all the experiments, we used stratified 10-fold cross-validation with 10 iterations, with tuning sets. We compared the results using a two-tailed corrected paired t-test, with \(p=0.05\). The best models found with the internal tuning were then applied to the test sets. We report the average number of Correctly Classified Instances (CCI), sensitivity and specificity, calculated according to what is discussed by Forman and Scholz\([5]\).

For the second goal, we focused on exploring the data trying to discover new knowledge. In WEKA, we used association rule mining, feature selection and used classifiers that produce interpretable results (e.g., J48). When trying to find relations among attributes, in the WEKA system, we tested all possible combinations of “Attribute Evaluator” and “Search Method”. The most frequent ranked attributes were used again to further filter and select the best attributes. Interestingly, the attribute Murmur was highly ranked. This step was done using 10-fold cross-validation and in each run, we selected the attributes that were most frequently selected in the 10 folds. We also used Aleph\([6]\), an inductive logic programming system that produces human-readable first-order rules. Experiments with Aleph were performed over the entire dataset.

3 Results
The best models found in the internal 10-fold cross-validation (tuning) were applied to the test sets. When predicting CardiacPathology, with the attribute Murmur, in seven folds, the best classifier was Grading, and...
in the remaining folds, SMO. The overall performance on the tuning and test sets are shown in Table 1. The results on the tuning sets are statistically better than ZeroR, which we use as the reference classifier.

We repeated the same experiment, removing the attribute Murmur from our dataset (Table 2). For this experiment, the best results were always obtained with a Naïve Bayesian network classifier in all folds. The accuracy (CCI) is statistically worse than the accuracy achieved by the classifier that uses Murmur, and not statistically different from the reference classifier ZeroR. Not every murmur is related to a pathology, but in our dataset, the attribute Murmur seems to be very important to predict it. These results suggest that, if possible, the attribute Murmur should always be annotated. If not, with only the attributes we have, our classifier would have a poor performance. One alternative would be to extract the attribute Murmur from the wave sound through signal processing, but this is an open research issue and could be a challenging task. Concluding, the attribute Murmur needs to be annotated and, according to these experiments, is crucial to obtain a classifier that can predict cardiac pathology with good sensitivity and specificity.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Tuning</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCI (%)</td>
<td>91.56</td>
<td>90.53</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.72</td>
<td>0.70</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.98</td>
<td>0.97</td>
</tr>
</tbody>
</table>

**Table 1 – Results with Murmur**

When doing feature selection, the relevant attributes chosen by all algorithms were: BMI, def, Age, def, Sex, SystolicSystemicPressure, def, DiastolicSystemicPressure, def, Hypertension, Murmur, Grading, S2Status, IfAbnormal, PulmonaryComponent, CardiacPathology and CardiacPathologyType, which coincide with the attributes we used for classification.

The HotSpot algorithm correlated the CardiacPathology attribute (class variable) with BMI (these results were also obtained using a reduced set of patients [7], where the relationship was for height and weight, which are the basic attributes used for computing the BMI). Removing the attribute Murmur maintains this relationship, but BMI is replaced by Sex. Similarly, when trying to discover the best attributes to predict the class variable (CardiacPathology), all algorithms select CardiacPathologyType and Murmur. In the absence of either or both of these attributes, S2Status, IfAbnormal and SystolicSystemicPressure, def are selected. These are all clinically relevant variables related to cardiopathies.

When learning first-order rules, we found an intriguing rule. This rule says that if a child has a systolic murmur and a high BMI (Body Mass Index), it is very likely that the child has a pathology. BMI in children is rarely related to cardiac pathologies according to most specialists. This rule may open a new stream of research into this relationship in clinical practice. This rule does not contain the sex and age of the child and this omission needs to be further investigated. The rule holds for 6 out of the 40 patients with a cardiac pathology, and does not apply to any healthy patient (129). This finding has been discussed in other work. Daniels et al. [8] mention that classic signs and symptoms of heart failure are not always present in obese patients, whose body habitus may mask signs of edema and may muffle the heart and lung sounds during auscultation. In their study, patients with high BMI were less likely to have documented murmurs. In our dataset, the opposite seems to be true, since we had 6 patients with annotated murmurs and a high BMI.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Tuning</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCI (%)</td>
<td>79.37</td>
<td>79.29</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

**Table 2 – Results without Murmur**

When learning rules, we uncovered an intriguing one that relates BMI with Murmur and CardiacPathology. Usually, BMI is not considered relevant to predict cardiac pathologies in children.

When learning classifiers, results show that we can train a classifier close to a specialist with a performance of 90.5%, sensitivity of 0.70 and specificity of 0.97 to predict pathologies on unseen cases. The area under the ROC curve was 0.83. We consider these results very promising and have been working to acquire more data to improve the classification task.

The work developed in the context of DigiScope has been paving the way to a new vision of cardiology. One important stream to follow is that of education in cardiology. The sounds recorded by DigiScope are being used to train novices to identify basic and more challenging cardiopathies. A second and also challenging path is to process the signal and extract important indicators such as amplitude and distance between signals (A1, P1, A2, P2 etc). With these indicators, we believe it will be possible to learn more effective classifiers. Our final goal is to have an integrated tool, capable of online predicting the cardiac pathologies and recommending additional screening. Ongoing work is being done to acquire data from more patients. We also have been working with data from adults and from pregnant women, which poses new challenges on the auscultations and annotations of the sounds.

**Acknowledgements**

This work has been supported by the project DigiScope (PTDC/EIA-CCO/10084/2008), the ERDF/COMPETE Programme and by FCT within project FCOMP-01-0124-FEDER-022701.

**Note**

This work was published in the proceedings of the 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2012) and subject of an oral presentation in Rome – June 2012 [9].

**References**


Fast performance 3D object recognition: Dealing with rotationally symmetric objects

Rui Pimentel de Figueiredo
ruifigueiredo@isr.ist.utl.pt
Plinio Moreno
plnio@isr.ist.utl.pt
Alexandre Bernardino
alex@isr.ist.utl.pt

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisboa, PT

Abstract

In this paper we extend a recent approach for 3D object recognition in order to deal with rotationally symmetric objects. These are frequent in daily environments and can be recognized with significant computational savings if symmetrical features are jointly represented. We show improvements up to 120x with respect to state-of-the-art methods.

1 Introduction

In recent past, Drost et al. [3] proposed an approach which extracts descriptor from a given object model, using point pair features [6], which encode the geometric relation between oriented point pairs. The matching process is done locally using an efficient voting scheme similar to the Generalized Hough Transform (GHT) [1]. Their method is robust to sensor noise and outperforms other feature-based state-of-the-art methods like Spin Images [4] and Tensors [5], both in terms of robustness to occlusion and clutter and in terms of computational speed.

In this paper we introduce an important extension to [3] for dealing efficiently with rotationally symmetric objects, which are common in many environments (e.g. kitchenware objects like cups, glasses, cans, plates). We drastically reduce the computational effort of [3] when dealing with this kind of objects.

2 Method Overview

Each model is represented by a set of points and associated surface normals, i.e. surfels [6]. Let \( M \) be the set of all model surfels, \( M = \{ s_i^m, i = 1..N \} \) – upper indices \( m \) and \( s \) will be further used to distinguish model from scene, respectively. An object description suitable for object recognition and pose estimation is created through the analysis of all possible permutations of surfel pairs. Let \( A \) be the set of all surfel pairs, \( A = \{ \{ s_i^m, s_j^n \}, r \neq t \} \), which has cardinality \( |A| = N \times (N-1) \).

2.1 Model Description

For each surfel pair \( (s_r, s_t) \), we compute a descriptive 4-element feature vector as illustrated in figure 1. This can be formally described by the following expression:

\[
F(s_r, s_t) = (f_1, f_2, f_3, f_4) = (||d||, \angle(n_r, d), \angle(n_t, d), \angle(n_r, n_t))
\]

The data structure used to represent the model description is a hash table for quick retrieval, in which the key value is given by the discrete point pair feature while the mapped value is the respective surfel pair. Since one key could be associated with several model surfel pairs, each slot of the hash table contains a list of surfel pairs with similar discrete feature.

Figure 1: Point pair feature descriptor

2.2 Pose Estimation

A set of reference surfel pairs on the scene \( R_s \subset S \) – where \( S \) is the set of all scene surfel pairs, \( S = \{ s_i^j, i = 1..N \} \) – is randomly chosen and each of them is paired with all the other surfels on the scene. For each scene surfel pair \( (s_i^l, s_j^k) \in S^2 \) we compute a point pair feature \( F(s_i^l, s_j^k) \) and then, using the extracted feature, we obtain a set of model surfel pairs whose feature is similar to it. From every match between a scene surfel pair \( (s_i^l, s_j^k) \in S^2 \) and a model surfel pair \( (s_i^m, s_j^n) \in M^2 \), we are able to extract the rigid transformation that aligns the matched model with the scene. This is done by first computing the transformations \( T_{m \rightarrow s} \) and \( T_{s \rightarrow m} \) that align \( s_i^m \) and \( s_j^n \), respectively, to the object reference coordinate frame \( x \) axis, and secondly the rotation \( a \) around the \( z \) axis that aligns \( p_i^m \) with \( p_i^l \). The final transformation that aligns the model with the scene is then computed considering the ensuing expression:

\[
T_{m \rightarrow s} = T_{s \rightarrow m}^{-1} R(a) T_{m \rightarrow s}
\]
The transformations $T_{m \rightarrow g}$ and $T_{r \rightarrow g}$ translate $p_m^a$ and $p_r^a$, respectively, to the reference coordinate frame origin and rotates their normals $n_m^a$ and $n_r^a$ onto the x axis. After applying these two transformations, $p_m^a$ and $p_r^a$ are still misaligned. The transformation $R(\alpha)$ applies the final rotation needed to align these two points.

The transformation expressed in eq. (4) can be parametrized by a surflet on the model and a rotation angle $\alpha$. In [3], this pair $(s^e, \alpha)$ is mentioned as the local coordinates of the model with respect to reference point $s^e$.

### 2.2.1 Voting Scheme

This method uses a voting scheme similar to the GHT for pose estimation. For each scene reference surflet, a two-dimensional accumulator array that represents the discrete space of local coordinates is created. The number of rows, $N_m$, is the same as the number of model sample surflets $|M|$, and the number of columns $N_{angle}$ is equal to the number of sample steps of the rotation angle $\alpha$.

The voting procedure goes as follows: considering a given reference surflet $s^d$ on the scene surface $S$, we pair it with every other surflet $s^f \in S$. For each resulting surflet pair we search on the model surface for similar surflet pairs, with the aim of finding where it might be in the model. This is done by querying the model descriptor for surflet pairs with similar feature. The computed feature $F(s^f, s^f)$ is used as an index to the model hash table and a list of matched surflet pairs, with similar feature, is returned. For each match $(s^f, s^f, s^f)$ the rotation angle $\alpha$ is computed and a vote is placed in the accumulator array by incrementing the position correspondent to the local coordinates $(s^e, \alpha)$, by the weight of the matched model surflet pair. After pairing $s^d$ with all $s^f$, the highest peak – i.e. the position with more votes – in the accumulator corresponds to the optimal local coordinate. In the end, all retrieved pose hypotheses whose position and orientation do not differ more than a predefined threshold are clustered together.

To deal with symmetry, before clustering, we collapse all redundant hypotheses to a single pose. This additional step removes the rotational component around the object axis of symmetry, i.e. yaw, ensuring that all redundant poses are gathered in the same cluster, therefore allocating less resources and reducing the number of computations.

### 3 Results

To evaluate the quality of the poses recovered by the algorithm and its runtime performance, we generated 200 synthetic scenes containing a single instance of a symmetric cup. Each scene was then corrupted by different levels of additive Gaussian noise, with standard deviation proportional to the model size. By using synthetically generated scenes, we were able to compare the algorithm pose results with a known ground truth. During recognition we chose 5% of the scene points as reference points. A higher percentage would increase the robustness to noise but also the recognition runtime. A recovered pose was considered to be correct if the error relative to the ground truth pose was smaller than diameter($M$) / 10 for the position and $12^\circ$ for the orientation. Thresholds from expressions (2) and (3) were set to $12^\circ$ and diameter($M$) / 40, respectively. The method was implemented in C++ and the experiment was run on a single core of a dual-core 2.6 GHz computer with 4GB of RAM. We were able to discard near 93% surflet pairs during the creation of the model description, and reduce the number of computations during pose recognition. As shown in figure 4, the recognition rate drops slightly for high levels of noise due to sampling effects, but the recognition time performance increases significantly. For $|S| \approx 5000$, our method achieves a recognition time 120 times faster than [3].

![Figure 3: Example of surflet pairs with similar feature stored in the same slot of the hash table, during the creation of the object model description.](image)

![Figure 4: Comparison results of our method against the method of Drost et al., with $|R^a| = 0.05 |S|$ reference points.](image)

**References**


Comparative Study of Inverse and Forward Sensor Models in Occupancy Grid Mapping Using Sonars

João Carvalho
joaopscarvalho@ist.utl.pt
Rodrigo Ventura
rodrigo.ventura@isr.ist.utl.pt

Abstract

Building occupancy grid maps with sonar sensors is a challenging task due to angular uncertainty, specular reflections and crosstalk. This paper presents a qualitative comparison of two probabilistic approaches to the robotic mapping — inverse and forward sensor models — and proposes a different formulation to the latter. The inverse one assumes independence of the cells while the forward is formulated as a maximum likelihood problem over a binary grid.

1 Introduction

One of the most used map representation in robotics is the occupancy grid map (OccGrid map), which aims to geometrically represent the environment through a grid discretization of the space. To build this maps, one commonly used sensor is the sonar. Sonars are cheap and allow the construction of maps even with a low number of sensors. Despite these advantages, sonars suffer from angular uncertainty, specular reflections and crosstalk between each other, causing erroneous and conflicting measurements [5].

The typically used method to build OccGrid maps is the one proposed by Elfes, making use of inverse sensor models [2]. In this approach, the occupancy of each cell is computed disregarding the rest of the map. A different approach was proposed by Thrun, using forward sensor models [4]. This method approaches the mapping problem in the high-dimensional space of all maps, trying to solve erroneous and conflicting sonar measurements which affect Elfes’s method results. This paper aims to brieﬂy compare this two approaches and propose a slightly different formulation to the latter.

2 OccGrid Maps with Inverse Sensor Models

In this approach, the mapping problem is treated inversely to how sonar data is generated, being formulated as

\[ p(M|z_{1:T},x_{1:T}) \]

where \( M \) represents the complete map, \( z_{1:T} \) represents the complete set of measurements and \( x_{1:T} \) are the corresponding poses. This is the denominated inverse sensor model.

To simplify the mapping problem, it is assumed that the occupancy of a given cell is not important to the computation of the occupancy of its neighbour cells, i.e., cells are conditionally independent given measurements and the robot trajectory, transforming the mapping problem into a binary estimation problem,

\[ p(m_i|z_{1:T},x_{1:T}) \]

where \( m_i \) is an individual cell of the complete map. A second assumption is that the static world assumption, considering a measurement \( t \) conditionally independent from the previous measurements given the map knowledge. This is a common assumption in mapping but given the decomposition into a binary problem this becomes a much stronger and also incorrect assumption, since it considers conditional independence given only a map cell and not the complete map. Additionally, the pose in the instant \( t \) is independent from the poses in previous instant times. So, for time \( t \):

\[ p(z_i|z_{1:t-1},x_{1:t},m_i) = p(z_i|m_i,x_i) \]

Given these assumptions and applying the Bayes rule to (2), we have:

\[ p(m_i|z_{1:T},x_{1:T}) = \frac{p(m_i|z_{1:T})p(z_{1:T}|m_{1:i-1},x_{1:T-1})}{p(m_i|z_{1:T})} \]

As is common practice, we will compute the log odds of this probability instead of the probability itself:

\[ \log p_{i} = \log \frac{p(m_i|z_{1:T})}{1 - p(m_i|z_{1:T})} - \log \frac{p(m_i)}{1 - p(m_i)} + \log p_{i-1} \]

where \( p_{i} \) represents \( \log \frac{p(m_i|z_{1:T})}{1 - p(m_i|z_{1:T})} \). The term \( p_{i-1} \) equals \( \log \frac{p(m_i)}{1 - p(m_i)} \) when \( t = 1 \). The probability \( p(m_i) \) is the prior of occupancy of the cell \( i \) of the map. A typical and simple approach is to model the posterior not as a fixed functional form but by a finite number of values which roughly approximate the posterior [5]. For the cells at distances between 0 and the neighbourhood of the measurement the occupancy probability has a low value, in the neighbourhood it has a high value and 0.5 beyond.

Making use of (5) the log-odds occupancy representation can be easily computed for each cell that falls into the coverage cone of the sonar measurements. So, finally, the desired occupancy probability of the cells can be recovered through:

\[ p(m_i|z_{1:T},x_{1:T}) = 1 - \frac{1}{1 + e^{-\log p_{i}}} \]

In the implementation, and in order to make it more robust to specular reflections, it was given less weight to larger measurements. So in (5), the term \( \log p_{i-1} \) comes multiplied by a variable, restricted between 0 and 1, that is inversely proportional to the sonar measurement \( z_i \). The final map is obtained after submitting each cell to a threshold. If the probability of occupancy is inferior to the threshold, the cell is considered unoccupied, being considered occupied otherwise.

3 OccGrid Maps with Forward Sensor Models and 1D Clustering

Static world assumption is also made in this approach but, in order to address the listed sonar problems, it does not make a map decomposition, dealing with mapping problem in its complete state space. Additionally it uses forward sensor models, being able to make use more complex sensor models. The forward mapping approach is modelled as a likelihood

\[ p(z_{1:T}|M,x_{1:T}) \]

This is a generative model, being formulated as the phenomenon happens; given the world (represented by the map \( M \)) and a given set of poses, a particular set of sonar reading is generated. The goal is to maximize (7), iteratively adjusting \( M \) till no better model is found. This problem can then be formulated as a maximum likelihood estimation problem.

Rather then assuming that all measurements are caused by an obstacle, three possible cases of beam reflection are considered, maximum reading, random and non-random. A non-random measurement is caused by an obstacle in the sonar beam. A maximum value reading happens due to angular uncertainty, specular reflections and crosstalk between each other, causing erroneous and conflicting measurements.

For the measurement with index \( t \), consider \( K_t \) to be the number of obstacle present in the sonar cone, \( d_k \) the distance from the \( k \)’th obstacle in the cone and \( D_t \) the set of obstacle distances in ascending order. The model (7) is deﬁned as the combination of the models of each possible cause. Consider the binary variables \( c_{i,k} = c_{i,k}, c_{i,0} \), restricted to:

\[ c_{i,k} + \sum_{k=0}^{K_t} c_{i,k} = 1 \]

These variables are equal to 1 when the measurement is random, caused by obstacle \( k \) or equal to the maximum range, respectively. The random case is modelled as a uniform distribution in the entire sonar range, since the reading could have been caused in any part of the sonar cone. When the beam is reﬂected by an obstacle, it is considered that it is affected by

\[ \log p_{i} = \log \frac{p(m_i|z_{1:T})}{1 - p(m_i|z_{1:T})} - \log \frac{p(m_i)}{1 - p(m_i)} + \log p_{i-1} \]

\[ p(m_i|z_{1:T},x_{1:T}) = 1 - \frac{1}{1 + e^{-\log p_{i}}} \]
Since there is no prior knowledge of the measurement’s cause we define the posterior probability

\[ p(z|M,x) = p(z|x,M) = \sum_{c} p(z|x,c) p(c|M) \]  

To maximize (17), a variation of Dempster’s Expectation-Maximization algorithm is used, where \( c_i \) is a vector of hidden variables [1]. No terms are discarded in (17), since any change in \( M \) might produce significant value variations in those terms. To find the map \( M \) that maximizes the likelihood, the occupancy of the cells that fall into the measurements cone is flipped and maintained if its new value increases the likelihood value. Given the discretization made, this results in a very greedy algorithm, in which the final result highly depends on the cell flipping order. Since it gave empirically good results, in this implementation we chose to first flip the cells closest to the measurement and progressively moving away. The Dirac delta function in (11) is implemented as a gaussian distribution with a very low variance.

4 Results

The robot used was the Pioneer P3-AT, equipped with SensComp 600 Series sonar sensors and a Sick LMS200 laser range finder. A ground truth map was built using the laser and the GMapping algorithm [3]. Figure 1(a). Since it is assumed that the robot’s pose is known, the GMapping’s pose estimative was assumed as the true pose. To build the maps two sonars were used, placed orthogonally to the robot’s movement and one on each side. The measurements were taken with the robot moving approximately at 0.6 m/s and measurements being taken with a 4 Hz frequency on a single lap to the corridor. Both approaches present an overall good representation of the environment. The forward method presents less artifact obstacles while fails to represent some obstacles in the corridor atrium, moreover it is prone to local maxima and is very computationally intensive.

5 Conclusion

This paper presented a comparison between OccGrid mapping using inverse and forward sensor models. The preliminary results showed that the latter solves some problems that affect the inverse but having as drawback the computational effort. Future work consists in presenting a more quantitative and systematic comparison and in studying better approaches using forward sensor models, to improve computational efficiency and to deal with outliers.

References

Abstract
Eye tracking systems are currently becoming more and more useful and practical in several areas, especially in Neuroscience, to assist in the detection and evaluation of cognitive disorders. This work proposes an autonomous method for evaluating situations of visual contact between two users in a frontal conversation. This evaluation is based on the gaze direction of both users, under free head movements, using a pair of calibrated cameras. The proposed stereo configuration is calibrated using images of planar mirror reflections. The experimental results showed that this system is able to detect up to 87% of real visual contact situations, using a gaze estimator with a mean absolute error of approximately 1.8° in the horizontal direction and 2.2° in the vertical direction.

1 Introduction
Most of the gaze tracking systems are based on Pupil Center Corneal Reflection (PCCR) methods. These products are quite accurate however the problem is that all of them require expensive hardware or artificial environments, becoming intrusive and uncomfortable to the user. Alternatively, many methods have been proposed to estimate eye gaze directly from the iris or pupil contours using ellipse fitting approaches. Some of those methods either only allow small head movements or are based in the distance from the iris center to a reference point (e.g. the eye corners), sometimes requiring a fixed face orientation or limiting the head movements to avoid face features occlusion. There are some appearance based approaches that showed very accurate results on gaze estimation, like [3]. However, they all propose an exhaustive training method (e.g. in [3] a training method of 33 samples and a video clip is proposed).

The methodology that is presented in this work is similar to the one proposed by Takahiro Ishikawa et al. [4]. The same 3D geometric eye model is adopted. However, the face features tracker used in this work is the same as the one developed by Martins and Batista in [5]. This work consists on creating an autonomous system that is able to evaluate if there is visual contact between two users in a frontal conversation, with no head movement limitations. Something like this requires an head tracker, an iris tracker, an head pose estimator, a gaze estimator and a special type of stereo calibration. In the end, a wide rectangle around the eyes of each user will be assumed as being the region of interest (ROI), in which the gaze of the opposite user will be intersected in order to find out whether there is visual contact or not.

2 Head and Eye Tracking
In order to accurately extract the eyes region, a model-based approach to detect 58 facial features is implemented. From the eye images, the irises centers will be estimated and with that it is then possible to estimate the gaze orientation. Allowing the user to freely move the head will add 6DOF to the problem, which must be considered to compute the gaze direction.

2.1 Active Appearance Model
An active appearance model (AAM) is a statistical based template matching method (used, in this case, to model faces) that operates on parametric models of shape and texture, where the variability of the models is captured from a representative training set that is built during an offline phase. More specifically, the AAM addressed here is a combination of two independent models: one to model shape and another one to model appearance (see [5] for a better understanding about these two models and about the AAM fitting process into an arbitrary image \(I(x)\)).

2.2 Head Pose Estimation
The POSIT algorithm [2] is a fast and accurate iterative method to find the 6DOF pose of a 3D model with respect to a camera. The method that was used in this work is proposed by Martins and Batista in [5]. With the one-to-one point correspondences between the AAM (2D points) and an anthropometric model of the head (3D points), the POSIT algorithm can be easily applied. The orientation of the estimated pose is represented in RPY (Roll, Pitch and Yaw) angles and the translation (with respect to the camera’s center) is given by \(\mathbf{T} = [T_x \ T_y \ T_z]^T\).

2.3 Eye Tracking
The gaze estimation is based on two points (for each eye). They are: the center of the eyeball (that will be explained further on) and the center of the iris. In order to detect circular contours, the Daugman’s integro-differential operator [1] was used, because it works with raw derivative information, with no need of threshold values. The iris contour is defined by its radius \(r\) and center \((x_0, y_0)\). These parameters can be determined by using the integro-differential operator

\[
\max_{(r,x_0,y_0)} G_\sigma(r) \ast \frac{\partial}{\partial r} \int_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds
\]

where \(I(x,y)\) is the eye image, \(G_\sigma(r)\) is a Gaussian smoothing function and \(s\) is the contour of the circle given by \(r, x_0, y_0\). This operator finds all circles in \(I(x,y)\), computing the sum of pixel values within each circle. Then the values of adjacent circles are compared and the returned contour is the circle with the maximum difference from its adjacent circles.

3 Gaze Estimation
In order to estimate the 3D eye gaze with just a single camera, it is necessary to define a 3D structure for the eyeball. The eye model that is adopted in this work is the same as in [4], and it is illustrated in fig. 1(a). From this model is important to extract the anatomical constants: \(R\) is the radius of the eyeball; \(T_x, T_y\) is the offset between the mid-point of the eye corners and the eyeball center; and \(L\) is the depth of the center \(o\) relatively to the plane containing the eye corners. These constants change from person to person, and therefore an unique offline calibration is required. It is a 3-point calibration where the user’s head is completely frontal (RPY = 0°). The offset \((T_x, T_y)\) is computed when the user is looking at the center of the camera, assuming that the image projection of \(o\) (on fig. 1(a)) coincides with the iris center \(p\). Using two other known points, the eyeball radius is estimated according to the method described in [6]. The depth \(L\) is related with \(R\) by a fixed value. Having these constants defined for
every user, the gaze direction can then be estimated by

\[
\sin \theta_i = \frac{p_i - o_i}{\sqrt{R^2 - (p_i - o_i)^2}}. \tag{2}
\]

The angles \(\theta_i\) and \(\theta_o\) represent the gaze orientation of the eye on the X-axis and Y-axis, respectively. The center of the eyeball \(o = (o_x, o_y)^T\) is where the head pose compensation will be applied. It is computed by applying two corrections on the mid-point \(m = (m_x, m_y)^T\)

\[
\begin{bmatrix} o_x \\ o_y \end{bmatrix} = \begin{bmatrix} m_x \\ m_y \end{bmatrix} + S \left( T_x \cos \phi_y + T_y \sin \phi_y \right) + S \left( \sin \phi_y \phi_x \right). \tag{3}
\]

In equation (3), \((\phi_x, \phi_y)\) is the head pose. The eyeball centre estimation requires the use of the scale factor \(S\), because for each frame the scale of the face changes, and the anatomical constants were only calculated for an unique \(S\). This scale is computed using the foreshorten-corrected distance between the eye corners [4].

4 Visual Contact Evaluation

In the final phase, the goal is to know whether one person is looking or not to the other person’s eyes. To achieve this it is necessary to set up a specific stereo camera configuration.

4.1 Stereo Calibration

Two cameras were attached, side by side, to a framework and brought into the middle of the conversation space (like fig. 1(b) illustrates). Notice that both cameras are only aware of the position (or even existence) of the person that lies inside their FOV. The drawback of this configuration is that the cameras don’t have a common FOV, what makes it impossible for the person that lies inside their FOV. The drawback of this configuration is that the cameras are only aware of the position (or even existence) of the person that lies inside their FOV. The drawback of this configuration is that the cameras don’t have a common FOV, what makes it impossible for the person that lies inside their FOV.

The problem of eye gaze estimation is geometric problem where the gaze line (for both eyes) is intersected with the pattern position. Basically, one of the cameras will be calibrated using this method while the other camera has the calibration pattern inside its FOV (this camera is calibrated by using Jean-Yves Bouguet’s toolbox). After performing both individual calibrations for each camera, there will be two different transformations that relate both cameras to the same pattern. The relation between the two cameras will then be given by

\[
C_{\text{Cam}0} T_{\text{Cam}1} = C_{\text{Cam}0} T_{\text{pattern}} T_{\text{pattern}} T_{\text{Cam}1}. \tag{4}
\]

4.2 Visual Contact Classification

To know whether there’s visual contact or not, the eye gaze of one user will be intersected with the eyes region of the opposite one. This region of interest (ROI) is defined by the head pose and two specific eyelid points which 3D position is extracted in every frame. This ROI can be projected from one camera coordinate system to the other camera coordinate system by using the transformation that was obtained in eq. (4). At this point, both cameras are aware of the gaze direction of one user and the eyes regions of the other user. The visual contact classification is now just a geometric problem where the gaze line (for both eyes) is intersected with the eyes region. If the point of focus of the gaze lines, when intersected with the eyes plane, lies inside the eyes region of the opposite user, then there is visual contact.

5 Experimental Results

After creating the offline shape model, the AAM will accurately track the user’s face. The Daugman’s algorithm reported very good results even under bad lighting conditions and with large iris occlusions, like fig. 2 illustrates. To test the gaze direction error, a white board with black points was placed behind the camera and the user was asked to sit more than 1 meters away from it and look at the points sequentially. Two experimental results, in different conditions, are shown in figures 3(a) and 3(b).

Finally, in a conversation process, one of the users was asked to look at a specific point in the opposite user’s head (in the middle of the eyes). The graphs from fig. 3(c) show the comparison between the position of the real point and the measured point of focus. To deduce the mean error on the gaze estimation, an absolute mean error was computed for every point of fig. 3(b) (the worst scenario), and then a weighted average of those absolute mean errors was calculated, resulting on the final absolute mean error of the gaze estimator.

6 Conclusion

With the experimental results it is possible to say that the proposed method has an absolute mean error on the gaze estimation of approximately 1.8° on the horizontal direction and 2.22° on the vertical direction. The gaze errors that were noticed in the experimental results were never higher than 4.7°. On the process of conversation, the system will be affected not only by the gaze direction errors but also by the stereo calibration errors. The final visual contact evaluator proved accuracy during a regular conversation, detecting up to 85% of real situations where there is visual contact.

It is then appropriate to admit that with some improvements in the head pose estimation (using a non-rigid head model) and in the anatomical constants calibration, it is possible to obtain an higher detection rate, making this system suitable mainly for medical applications such as attention and cognitive disorders auxiliary evaluation.

References

Hand Pose Estimation using a Top-Down Bottom-Up based Approach

Davide Periquito
davide.periquito@gmail.com
Jacinto C. Nascimento
jan@isr.ist.utl.pt
Alexandre Bernardino
alex@isr.ist.utl.pt
João Sequeira
jseq@isr.ist.utl.pt

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisboa, Portugal

Abstract

Most existing systems require a manual/supervised initialization in order to perform hand tracking in a video sequence. Furthermore, fully automatic initialization is a very challenging problem due to the difficulty in identifying landmarks on individual fingers and reconstructing the hand pose from their positions. In this paper, we propose an appearance-based approach relying on matching hand silhouettes to a previously trained database. A dense sampling of the hand appearance space is obtained through a simulation environment and the corresponding silhouettes stored in a database. The acquired silhouettes are efficiently retrieved on-line from the database using a framework that combines a bottom-up and top-down processes. We testify the performance of the proposed approach in a series of simulations, evaluating both the influence of the bottom-up and top-down processes with respect to estimation error and computational cost.

1 Introduction

Human Computer Interaction (HCI) has received considerable attention in the Computer vision community. However, to become practical, several issues must be accomplished, among which: self-start; robust tracking throughout the sequences; robustness against drift and occlusions; computational efficiency; and avoid background subtraction. In this paper, we will focus on the first issue, i.e. automatic start/restart of the tracker, being one of the most challenging problems in HCI.

In this paper we present a method to address this problem by using an algorithm based on a bottom-up and a top-down processes. A very quick bottom-up approach filters out the pose candidates set, so that the more computational intensive top-down process only has to evaluate a reduced set of good candidates. The algorithm initially builds a training set with known postures. In run time the observed image is matched against the trained set of hypotheses using two matching metrics with different computational costs and precisions: first, with the geometric moments (bottom-up) to perform a quick filtering of the training set, and then with the Jaccard distance (top-down) to achieve a more reliable posture hypothesis.

2 Algorithm

This section describes the main steps of the proposed framework which includes: (i) generation of the training images, (ii) segmentation and localization, and (iii) pose estimation.

2.1 Training stage

The hand pose hypotheses are generated using the OpenRAVE. More specifically, we place a virtual camera on the simulated model pointing at the 3D hand model. By moving the pose around at a constant distance to the hand we create a virtual sphere path. To represent the orientation of the camera we use a quaternion representation. Uniform samples on the orientation space are generated by drawing quaternions from a Gaussian distribution. For each sample a difference of 5° (degrees) is guaranteed in the generation process. The camera rotation matrix is given as in [5]. These steps are resumed in the next [1-2] items.

1. Computation of the quaternions to generate the training hand pose hypotheses images. A total of 23900 images are used,

2. Hand pose hypotheses are then generated in the OpenRAVE simulator [2] with an existing humanoid 3D model. A total of 23500 images are used for training and 400 used as poses to test the framework,

3. The images are segmented (i.e. the silhouettes or contours are obtained) and corrected in perspective to simulate frontal views,

4. The geometric moments of the contours are computed,

5. The silhouettes are stored in a database, together with both the binary masks and the geometric moments. Also the ground truth poses (i.e. quaternions) are stored.

The above items [1-5], are accomplished off-line. The following step is the on-line test stage. Here, we perform the matching between the acquired hand silhouette and the pre-trained database of canonical pose hypotheses. Still on the fly, each acquired image silhouette is also pre-processed as in the training stage comprising color segmentation, perspective correction and binarization. A total of 400 images are used for testing. First, the geometric moments (bottom-up) of the newly acquired mask are used to rank the training set in descending order of match quality. We filter the top 1000 hypotheses candidates that are the output of the bottom-up step of the framework. Finally the top-down procedure which allows a more precise match using the Hammoude metric [4] also known as the Jaccard distance [3]. The top-down is applied over the top ranked candidates in order to provide for a final decision.

2.2 Segmentation and localization

For hand segmentation, we follow a simple approach. We use the HSV color space that allows better luminosity invariance. Then, a Histogram Backprojection algorithm is used [6] for segmenting the hand. This results in a histogram of the likelihood of each pixel voting to the hand localization. After the filtering process the result is a segmented hand. Some noise may be present in the segmentation. Thus, we make some image processing, by filling the holes inside the hand and removing border objects. In this way, we are able to obtain a clean binary image with a segmented hand. This method is identical for the training set images and for the test images from which we want to determine the hand position.

For better matching, the hand centroid \((x_0, y_0)\) is placed in the center of the image, though for this procedure the hand has to be rotated according to the displacement made before. So, the Homography equations for projecting and rotating points in an equivalent pan-tilt camera are, according to [1]:

\[
\begin{align*}
x_1 &= \frac{cr s_p + cp x_0 - s t p y_0}{cr s_p - sp x_0 - st c y_0}, \\
y_1 &= \frac{st + c r y_0}{cr s_p - sp x_0 - st c y_0}
\end{align*}
\]

where \(c_p, s_p, c_r, s_t\) stand for \(\cos(p), \sin(p), \cos(t), \sin(t)\), respectively, \((x_1, y_1)\) represent the pixels after the rotation, \(p\) and \(t\) are the equivalent pan-tilt camera angles, meaning that the previous \((x_0, y_0)\) are now centered in the camera. To compute the pan and tilt \((p, t)\) angles the translation of the image must be known, so:

\[
p = \arctan(x_1), \quad t = \arctan(y_1/c_p)
\]

ending with a segmented hand centered with the camera and projected according to the movement made. These changes of perspective introduce error in the appraisal though it is still acceptable.

Since we are working in a 2D image plane the Z coordinate is used for area normalization for the future matching metrics.
2.3 Pose estimation, Bottom-up (Geometric moments) and Top-down (Hammoude/Jaccard distance)

The two main ingredients of the pose estimation step are the bottom-up and top-down processes. The first comprises the computation of the geometric moments for the training set. The second consists on the computation of Hammoude metric that provides high accuracy in the pose estimation.

The bottom-up process is characterized by the use of the geometric moments to obtain the posture and shape of the hand, that are invariant to position and scale. This is achieved by the following two steps: (i) centering both of the geometric moment and the hand; (ii) centering the hand area normalization

\[
u_{pq} = \frac{\sum_{x,y}(x-x_0)(y-y_0)^2 I(x,y)}{M^2_{pq}}
\]  

(3)

where \(u_{pq}\) is the moment of order \(p+q\), \(M_{pq}\) is the hand area and \(I(x,y)\) is the image pixel. From our experiments, we concluded that it is essential to keep the moments of order higher than 4th, since the higher the order the more discriminative characteristics we get. In contrast, lower orders describe the hand position and area, which we want to be invariant.

To obtain the matching distance between trained and observed images, a Mahalanobis based distance is used

\[d = \sum_{p,q} (\hat{r}_{pq} - r_{pq})^2 \text{var}(n_{pq})^{-1}
\]  

(4)

where \(\hat{r}_{pq}\) is the moment calculated in an observed image, \(n_{pq}\) is the moment trained in the training set hypotheses and \(\text{var}(n_{pq})\) is the variance of the moment in the training set. By minimizing the function we have the most likely hypothesis.

For the top-down we use the so-called Hammoude metric [3,4], that allows the matching between the observed silhouettes and the one in the database and it is defined by

\[d_{HMD}(y_1,y_2) = \frac{\#((R_{y_1} \cup R_{y_2}) - (R_{y_1} \cap R_{y_2}))}{\#(R_{y_1} \cup R_{y_2})}
\]  

(5)

where \(R_{y}\) represents the image region delimited by the contour \(y\) (similarly for \(R_{x}\)), and \(\sharp\) denotes the number of pixels within the region by the expression in parenthesis. This is further converted in to a likelihood of each hypothesis by defining the probability \(p(y_1|y_2) = 1 - d_{HMD}(y_1,y_2)\).

3 Experimental results

In the experimental evaluation, we first study how to select the proper number of the geometric moments candidates to be used in the bottom-up. The experiment comprises the images generation as described in Section 2.1. To achieve the previously specified goal, we vary the number of output images in the bottom-up procedure. The geometric moments are varied in the interval range \(\mathcal{R} = [10, 100, 1000, 10000, 23500]\), and for each candidates number in the range \(\mathcal{R}\), we perform the hand pose estimation by using the top-down approach. The error metric used, is the orientation error defined as

\[\varepsilon = 2\arccos(p \cdot q)
\]  

(6)

where \(p \cdot q\) stands for the inner product between the quaternions of the two images. The error in eq. (6) is computed between the observed (test) image and the top rank hypothesis (in the training set). Finally, the average of the orientation error \(\varepsilon_{AV}\) is taken to assess the overall performance over the test set.

Table 1 establish a relation between the average of the orientation error \(\varepsilon_{AV}\) (in degrees) and the time consumed to assess the pose. We see a soft spot when using 1000 candidate moments, where the orientation error and the time spent are minimized. For online applications the time spent is of crucial importance, so in order to achieve such performance the geometric moments are a useful tool to efficiently reduce the training set.

Table 1: Mean and standard deviation (in parenthesis) of the orientation error \(\varepsilon_{AV}\) (in degrees) and time spent (\(^s\)-seconds, (ms)-milliseconds) for the hand pose estimation. The experiment is repeated for the top candidates moments defined in the range \(\mathcal{R}\).

<table>
<thead>
<tr>
<th># Cand. Mom.</th>
<th>Time</th>
<th>(\varepsilon_{AV})</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4.21 (s) (0.06 (ms))</td>
<td>7.34 (15.4)</td>
</tr>
<tr>
<td>100</td>
<td>5.99 (s) (0.98 (ms))</td>
<td>5.86 (3.06)</td>
</tr>
<tr>
<td>1000</td>
<td>11.90 (s) (2.74 (ms))</td>
<td>5.77 (3.00)</td>
</tr>
<tr>
<td>10000</td>
<td>69.07 (s) (5.33 (ms))</td>
<td>5.77 (3.00)</td>
</tr>
<tr>
<td>23500</td>
<td>122 (s) (8.11 (ms))</td>
<td>6.06 (3.50)</td>
</tr>
</tbody>
</table>

Figure 1: Accumulative rank using top 1000 candidate hypotheses.

Figure 2: Three snapshots of the sequence (top) poses recovered by the algorithm (bottom).

We tested our framework in real settings using a sequence of 50 frames. Fig. 2 shows some snapshots of the sequence as well as the recovered poses. It is concluded that a robustness regarding the pose is obtained (bottom row of the Fig. 2).

4 Conclusions

In this paper we proposed a 3D hand posture estimation framework. The architecture combines a bottom-up and top-down approaches, providing an efficient tool for hand orientation detection. The algorithm presented is twofold. First, the bottom-up allows for an efficient reduction over the training set, having a significant impact on computational time. Second, the use of the top-down provides an improvement estimation accuracy. Fusing these two processes, we can achieve a fast and accurate hand pose estimation in real environments.

Acknowledgement: This work was supported by the FCT project [PEst-OE/EEI/LA0009/2011]

References

Abstract

This paper describes a computer aided diagnosis system for the classification of cutaneous melanocytic lesions. Although the great majority of these lesions are benign, a small rate is melanoma, an aggressive form of skin cancer and the leading cause of death from skin disease.

Based on the characteristics highlighted by the ABCD rule of dermatology, the suspicious lesion is described by a set of characteristics that allow their classification as melanoma or non-melanoma.

1 Introduction

Most of the cutaneous melanocytic lesions are benign; however, some of them are melanomas, the most dangerous form of skin cancer. The worldwide incidence and mortality rates of melanoma are continuously increasing.

Dermoscopy is a non-invasive diagnostic technique for the in vivo observation of pigmented skin lesions (PSLs), allowing a better visualization of surface and subsurface structures. This diagnostic tool permits the visualization of morphological structures of the epidermis and the dermo-epidermal [1]. Several studies have shown that this method may improve diagnostic sensitivity by 20–30% compared with clinical diagnosis through the naked eye [2]. Because advanced skin cancers remain incurable, early detection and surgical excision is currently the only approach to reduce mortality, thus, the development of computer aided diagnosis systems for early detection of melanoma is crucial.

The ABCD rule, the three-point checklist, the seven-point checklist, and the Menzies’s method are the most relevant dermoscopic criteria and diagnostic algorithms proposed and analyzed in recent years by many publications for the classification of melanocytic lesions. The ABCD rule is a semi-quantitative analysis using the following criteria: asymmetric lesion (A), border irregularity (B), color (C) and differential structures (D) [3]. Through analysis of these attributes, TDS (Total Dermoscopy Score) can be calculated by equation (1):

\[ \text{TDS} = [(A \times 1.3) + (B \times 0.1) + (C \times 0.5) + (D \times 0.5)] \] (1)

The value of TDS is an indicator that helps specialists in the classification of the lesion. In this calculation the score of each feature is weighed according to its importance in the classification of the lesion.

The aim of this work is to transfer the asymmetry and color attributes (attributes with more weight), into automatically computed quantities, and use these parameters in order to automatically classify melanocytic lesions.

2 Segmentation of Skin Lesions

Lesion segmentation is an important step leading to the quantification of the geometric properties and color characteristics. In this case, segmentation is the process that isolates a pigmented skin region in dermoscopic images.

In the pre-processing phase the color images of the database are converted to gray images, equalized to enhance the contrast between the pigmented region and the skin. In order to remove parasite regions that introduce noise in lesion analysis, for example hair or isolated points, mathematical morphology is applied (Figure 1).

In a preliminary phase of the project the segmentation of the lesions was made through dynamic thresholding, using Otsu method and morphological transformations. This method proved to be effective in the great majority of the images of our experimental database, leading to their correct segmentation. However, for images where the lesion is poorly contrasted from the skin, lesions with smooth boundaries and in situations where the lesion consists of two or more separate regions, the method fails or does not isolate the entire lesion.

In a second phase we look for algorithms that increase the rate of useful segmented images. The binarization was made using Niblack’s algorithm that is a local thresholding method based on the calculus of the local mean and local standard deviation, followed by the algorithm proposed by Chan & Vese [4]. This algorithm is based on active contours and curve evolution using Mumford-Shad functional and level sets theory. It is a powerful algorithm that allows the correct segmentation of 58 images in a total of 63 images. Comparative results obtained with the two segmentation algorithms are shown in Figure 2.

3 Asymmetry of the Lesion

The asymmetry of the lesion was evaluated with regard to the border and color regularity with respect to two perpendicular axes that intersect at the centroid of the lesion. To simplify the analysis of the asymmetry, the image is rotated according to the angle of symmetry, calculated from geometrical moments, so that its major axis stays parallel to the X axis.

In accordance with the ABCD rule, the asymmetry of melanocytic lesions is categorized with a score from 0 to 2 points. In the case of a lesion completely symmetrical a score 0 is assigned, when asymmetric in relation to two axes, the score is 2 and when the lesion is asymmetric with respect to only one of the axes a score of 1 is assigned (Figure 3).
To evaluate the border asymmetry a binary mask is created based on the rotated image. This mask is divided in four symmetrical parts in relation to the X and Y axes (Figure 4 - Left). After the folding of the border the relative distances of the border are evaluated through the calculus of ratios of asymmetry relative to the X and Y axes:

\[
\text{Ratio}_x = \frac{|\text{Limit}_{\text{lower}} - \text{Limit}_{\text{upper}}|}{m_{00}} \tag{2}
\]

\[
\text{Ratio}_y = \frac{|\text{Limit}_{\text{right}} - \text{Limit}_{\text{left}}|}{m_{00}} \tag{3}
\]

where \(m_{00}\) is the total area of the segmented lesion.

Therefore, we developed an algorithm that identifies how many colors are present in a lesion calculating the color distance of all pixels in the lesion (\(d\)) with each of the colors reference, by equation 9. The algorithm developed quantifies the number of colors present in the lesion with a minimum value of 1 and a maximum of 6. This value is then inserted in the ABC rule with a weight of 0.5.

\[
d(x, y) = \sum_{k=1}^{3} |I^k(x, y) - I^k| \tag{9}
\]

In equation (9), \(I(x, y)\) represents the color (R,G,B) of pixel \((x,y)\) and \(I^k\) one of the colors reference of melanoma.

5 Classification

Our experimental dataset consists of 58 dermoscopic images, 24 with the presence of melanoma and 34 with non melanoma. All the images are classified by a specialist [5]. From the correct segmented lesions a set of features based on border and color symmetry and the number of colors present in the lesion are extracted, as proposed by the ABCD rule of dermoscopy.

The classification of the cutaneous lesion was performed using the state-of-the-art support vector machine (SVM) that belongs to a successful class of learning algorithms strictly connected with statistical learning theory. SVM is based on structural error minimization principle, which is a trade-off between the complexity of the learning machine model and its generalization capability.

The dataset is split into two different groups: the training and test set. In the training set a search for optimal parameters \((C, \sigma)\) of the RBF kernel is carried out using a grid search methodology through a \(k\)-fold cross validation. The optimized SVM model was evaluated in test set by the indices of accuracy, sensitivity and specificity.

6 Results and Discussion

From the total dataset, 50% was used for training and the remaining 50% for testing. The ability of the system to correctly classify new images is essential. So, in order to increase the classifier generalization, a 10-fold cross validation technique was performed in the training set. This methodology was repeated 100 times and the final mean accuracy, sensitivity and specificity were obtained.

The achieved results were an accuracy of 85.6% ± 0.09, a sensitivity of 80.6% ± 0.18 and a specificity of 85.4% ± 0.11. Analysing the set of extracted characteristics we can conclude that all the images of the dataset have the same range of color and a small diversity of colors. The maximum number of colors found was four and for example the white color was never detected. So we expect that by increasing the number of images in the dataset with more color diversity, the system performance can be improved, especially its sensitivity.

4 Color of the Lesion

Another feature of melanoma is the non-uniformity of color, while benign lesions are usually uniform brown color (Figure 5). The developed algorithm aims to quantify the number of reference colors proposed by the ABCD rule (black, dark and clear brown, white, blue, gray and red) present in the lesion with a minimum of one and maximum of six. We considered six RGB codes for each reference color to improve the flexibility of the algorithm and computed the distance for each channel, between the color of each pixel and the color that we intended to find.

Therefore, we developed an algorithm that identifies how many colors are present in a lesion calculating the color distance of all pixels in the lesion (\(d\)) with each of the colors reference, by equation 9. The algorithm developed quantifies the number of colors present in the lesion with a minimum value of 1 and a maximum of 6. This value is then inserted in the ABC rule with a weight of 0.5.

\[
d(x, y) = \sum_{k=1}^{3} |I^k(x, y) - I^k| \tag{9}
\]

In equation (9), \(I(x, y)\) represents the color (R,G,B) of pixel \((x,y)\) and \(I^k\) one of the colors reference of melanoma.

5 Classification

Our experimental dataset consists of 58 dermoscopic images, 24 with the presence of melanoma and 34 with non melanoma. All the images are classified by a specialist [5]. From the correct segmented lesions a set of features based on border and color symmetry and the number of colors present in the lesion are extracted, as proposed by the ABCD rule of dermoscopy.

The classification of the cutaneous lesion was performed using the state-of-the-art support vector machine (SVM) that belongs to a successful class of learning algorithms strictly connected with statistical learning theory. SVM is based on structural error minimization principle, which is a trade-off between the complexity of the learning machine model and its generalization capability.

The dataset is split into two different groups: the training and test set. In the training set a search for optimal parameters \((C, \sigma)\) of the RBF kernel is carried out using a grid search methodology through a \(k\)-fold cross validation. The optimized SVM model was evaluated in test set by the indices of accuracy, sensitivity and specificity.

6 Results and Discussion

From the total dataset, 50% was used for training and the remaining 50% for testing. The ability of the system to correctly classify new images is essential. So, in order to increase the classifier generalization, a 10-fold cross validation technique was performed in the training set. This methodology was repeated 100 times and the final mean accuracy, sensitivity and specificity were obtained.

The achieved results were an accuracy of 85.6% ± 0.09, a sensitivity of 80.6% ± 0.18 and a specificity of 85.4% ± 0.11. Analysing the set of extracted characteristics we can conclude that all the images of the dataset have the same range of color and a small diversity of colors. The maximum number of colors found was four and for example the white color was never detected. So we expect that by increasing the number of images in the dataset with more color diversity, the system performance can be improved, especially its sensitivity.

Referências


A Signal Acquisition and Processing Device to Assist Human Heart Sound-based Diagnosis

Jérôme Nameche,1,2,5
jnameche@gmail.com
Pedro Amaro3,4
amaro@isec.pt
Marco Silva,2,3
msilva@isec.pt
Fernando Ferreira2,3
fernando@isec.pt
Fernando Lopes5,4
flopes@isec.pt

Abstract

This paper presents a low-cost portable signal acquisition and processing device with application to the medical diagnosis of heart diseases through the time and frequency analysis of the human heart sound. Human heart diseases leave identifiable traces in the heart sound that can be isolated using signal processing techniques. A hardware device consisting on a microphone-adapted stethoscope, an acquisition and amplification circuit and a dsPIC-based processing unit was developed, that can be successfully used for this purpose. An example algorithm, focused on heart sound murmurs, that can classify the heart sound in four health classes, was implemented. The algorithm considers the patterns observed from a database of healthy and non-healthy hearts. This exploratory prototype was tested in a group of hospital patients proving to be especially adequate to assist in the diagnostic of valvular heart diseases. This type of low-cost and simple-to-use device can be of great interest in developing countries, where universal medical care is not available, providing a cost-effective, non-specialised identification of cases in urgent need of medical attention.

1 Introduction

Cardiovascular diseases relate to a class of diseases that involve the heart or the blood vessels, merging inter-related heart and vascular diseases [1]. Cardiovascular diseases have become one of the major health problems in developed and developing countries and remain the biggest cause of death worldwide. Causes can be congenital, infections, other diseases or a non-healthy life associated for example with the lack of healthy eating, lack of exercise, air pollution, excess alcohol and smoking. Not everyone has close access to a doctor that can analyze the heart sound or the money to afford a medical consultation. For efficient screening in developing countries, a single doctor would have to analyse an enormous number of heart sounds, what is impossible in practice. Therefore, the availability of a simple diagnostic assistance tool, that can be operated for instance by a nurse, can be an efficient solution for large scale screening, increasing the number of analysed patients and saving lives. In this paper we present a cost-effective, easy-to-use handheld device that can provide valuable assistance in early heart disease diagnostics. The device integrates a common stethoscope and can classify the heart sound in four classes: “Heart sounds very good”; “Heart sounds good”; “You should visit a doctor” and “You must visit a doctor”. “Imperfect acquisition” is also a possible output result.

2 Heart Diseases and Murmurs

Cardiovascular diseases in general and heart diseases in particular consist on a large set of complex interrelated conditions. Some prominent examples of heart diseases are Coronary Artery Disease, Cardiomyopathy, Hypertensive Heart Disease, Inflammatory Heart Disease and Valvular Heart Disease [1].

During the systolic and the diastolic phases of the cardiac cycle, audible sounds are produced due to the opening and closing of the heart valves, the flow of blood in the heart and vibration of the heart muscles. Heart diseases cause physical changes to the heart and blood vessel structure that can potentially be detected in the acquired heart sound. A typical phonocardiogram envelop from a healthy heart can be observed in Figure 1 where the heart anatomy is also presented.

Murmurs are the result of abnormal turbulent blood flow which produces a series of abnormal vibrations in the cardiac structure and that are not present when hearing a healthy heart [2]. Murmurs can be classified by their intensity, duration, timing, location, transmission and quality. They can also be described by their time-shape. Murmurs are especially observable in conditions associated with valvular diseases.

1 Haute Ecole Paul-Henri Spaak, ISIB Brussels, Belgium
2 Instituto Superior de Engenharia de Coimbra (ISEC), Instituto Politécnico de Coimbra
3 Instituto de Sistemas e Robótica, Universidade de Coimbra, Portugal
4 Instituto de Telecomunicações – Pólo de Coimbra, Universidade de Coimbra, Portugal

Figure 1: Heart anatomy and typical healthy phonocardiogram [3].

Figure 2: Phonocardiograms of normal and abnormal heart sounds. The names of the associated conditions are also indicated in this figure.
4 Application and Results

To validate the hardware and algorithm for the intended objective, the device was tested in three environments: using a source database testing for 8 healthy sounds and 16 disease sounds from [5]; using 12 healthy sounds acquired in the ISEC Campus; using 12 acquired sounds in hospital patients with diagnosed heart diseases. All healthy sounds resulted in “Heart sounds very good” while pre-diagnosed diseased hearts resulted in “You must visit a doctor”. In hospital patients, some acquisitions had to be repeated due to imperfect localization of the stethoscope. In Figure 5, four examples from the hospital patients test are presented for the “aortic stenosis” condition and where the relative high frequency contents can be observed. More extensive testing must be performed to tune the decision criteria to match the highest possible number of pre-confirmed high confidence medical diagnostics.

5 Discussion and Future Work

In this paper we presented a low-cost, easy-to-operate autonomous device that can assist heart diagnosis by identifying patients in need of medical attention. This ability can be very valuable in developing countries where widespread screening is not possible due to the scarcity of medical professionals, technical and economic resources.

To validate the developed hardware, an example algorithm, best suited to identify murmurs associated with valvular diseases, was implemented and successfully tested. The tests used a sound database and stethoscope acquired signals, both including healthy and pre-diagnosed diseased patients. A simple spectral criterion was used, which was found adequate for the health conditions that generate significant murmurs, as those associated with valvular heart diseases.

Future work will evolve in two directions. On the one hand, the algorithm for the portable device can capture more abnormal heart conditions and minimize the possibility of false negatives for the various conditions. This can be achieved exploring more frequency domain criteria, performing time-frequency analysis and training with a larger sound database. On the other hand, the study can evolve, not only to detect an abnormal condition, but also to contribute to an advanced system to assist detailed heart diagnosis and disease classification.

Acknowledgement

The authors would like to thank the support of Prof. Jacques Tichon and Dr. Eduardo Antunes from Santa Marta Hospital in Lisbon.

References

Applications for people with special needs using Kinect

Bruno Almeida (a21160603@alunos.isec.pt)
César Páris (cparis@isec.pt)
Anabela Gomes (anabela@isec.pt)
Joel Jegundo (a21171043@alunos.isec.pt)
Nuno Martins (ncmartin@isec.pt)
Álvaro Santos (ans@isec.pt)

Departamento de Engenharia Informática e de Sistemas
Instituto Superior de Engenharia de Coimbra - Instituto Politécnico de Coimbra

Abstract

This article describes a scientific project related to the Human-Computer Interaction (HCI), which was developed with the purpose of helping people with special needs, specifically with cerebral palsy.

The key points of this project are how a person’s body movements and his/her voice are detected by a device (in this case the Kinect sensor) and how this is translated into actions on the computer. All the applications should facilitate the access to learning and entertainment.

This project is divided into three areas:

- Help in education
- Accessibility
- Training and entertainment

To develop this project we rely on the help of APCC (a Portuguese acronym for Coimbra Cerebral Palsy Association).

1 Introduction

This paper describes a research work related with the creation of new human-machine interfaces related with accessibility concepts. The work was done using motion capture techniques based on the use of vision sensors and infrared sensors, both available on one system, called Kinect.

Kinect is a mixture of the Greek “kinesis” (that means moving or move) to “connect”. The name defines the sensor extremely well because the user is connected to a platform which enables him to go through the actions of his/her movements and gestures [1]. Kinect was developed by Rare (company that is part of Microsoft) and PrimeSense (which developed a system that can detect 3D object movements) [1]. In developing the Kinect, Microsoft demystified the paradigm of controllers to develop a new system of natural user interface that enables advanced human control, voice control, motion recognition among other things [2]. The Kinect uses multiple resources (image, sound, tilt motor, infrared, and depth) with high precision synchronization (in real time) in a single device.

Kinect has a parallel processing algorithm (programmed into the chip SoC – this designation is related to all elements of a computer that could be put in a integrated circuit), used for obtaining the depth map from the structured light received. [3] The images are aligned pixel by pixel so as to obtain more precise information of the sensors. These innovative features provide plenty of opportunities for interaction between services, applications, and users.

Through the camera, Kinect identifies the user by creating a three-dimensional image, to accurately identify the movements of the body. The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor that captures 3D video in any ambient lighting conditions. The range of the depth sensor is adjustable through software, Kinect is able to automatically calibrate the sensor based on the user and accommodating the presence of obstacles (furniture ...).

The software allows Kinect to recognize gestures, faces and voices. According to the manufacturers it is able to store information of up to six people simultaneously [4]. All this is a highly innovative combination of sensors and software that transform the user's body into control applications.

The paper is organized as follows: The next section explains each of the applications developed. The third section describes the actual development and the options taken. In the fourth section, some conclusions are drawn.

2 Developed applications

We developed applications in three different approaches. One approach was to help in education specifically in the area of mathematics, another concerns the computers accessibility, and the third deals with training and entertainment.

2.1 Help in education

This application aims to facilitate the learning process of basic math operations thinking mainly in children with cerebral palsy having difficulty in mobility.

We decided to widen the application interaction possibilities to its full extent, so that anyone could use it regardless of their restrictions. The user could control the application with his/her left or right hand, left or right foot or his/her head not limiting the application use.

We also have an option to use both hands simultaneously. The size of the buttons and click time is configurable to adjust the application to each specific user.

It also has an automatic mode that allows the result of each operation to be automatically verified at every insertion of a new digit in the result. If the result is correct a new math operation will be presented. Operations (addition, subtraction, multiplication and division) can also be configured by a user or a tutor through the specific menu. As can be seen in Figure 1 the operations are performed as in elementary school. In other words, the result is inserted from right to left.

The application has an initial calibration to verify the user interaction capabilities. If there are difficulties it is possible to change some configurations like the range of motion or the size of the buttons.

In order to familiarize the user with the application, and since it is an application focused mostly on children, the calibration is done through the use of appealing images, such as animals. This is done by touching animal's images that reproduce the sound of the specific animal.

On top of the window there's a circle which can be red, indicating that no user is being detected, or green when the user is being detected. This will help to verify that the reason why the application is not responding to gestures is the fact that no skeletal is being detected.

In order to facilitate the interaction of a second person (tutor) a set of keyboard shortcuts were created to interact with the application as well as interaction via computer mouse.

The application allows the Kinect sensor to be restarted, if for some reason it is needed, without having the need to restart the whole application.

2.2 Accessibility

Computers used by a person with reduced mobility are often a real problem, so we decided to develop an application so that the user can control the mouse cursor through gestures and using speech recognition.

The user has three options to perform a click:

- Only voice - one word for the click and another for double-click;
• A gesture and voice - a gesture and through a specific word the user can select whether to make the normal click or double-click;
• Two specific gestures - one to click and the other to double-click.

In Figure 2 we can see a user controlling the application with the left hand, and we can also check some configurable options - range of motion (Tamanho do Gesto), tilt angle (Ângulo) of the Kinect and what hand (Mão Esquerda / Mão Direita) will control the cursor.

2.3 Training and entertainment

Interaction with the Kinect is not a very complicated task, but it can improve through some training, so the more we use it the better we can control it. For this reason we decided to create a game, which is a fun way of exercising with Kinect. Other games can be developed for exercising the body in a pleasurable way. So, while the children are working they are having fun.

We created a series of colored boxes and through gestures the user can throw a ball affected by the gravity to hit the boxes, in the shortest time possible, forcing the user to move (Figure 3).

The game camera can be controlled by one hand and the opposite hand is used to throw the ball through one of two distinct gestures: one consists in raising the hand above the shoulder and the other consists in stretching the hand forward. Throwing the ball can also be controlled by the computer mouse. There are three stages in the game each having three difficulty levels (easy, normal and hard). The increase of difficulty consists in increasing the number of boxes as well as their weight. This game has been developed in XNA because it is an existing framework oriented to games, including a physics engine making the games more realistic.

3 Development and Usability

In this section we point out some of the problems inherent in applications that take advantage of Kinect capabilities and how we solve them.

3.1 How to use the application?

Most of the children that test our applications were in wheelchairs, making it difficult to detect the skeleton, creating an obvious need to find a solution to circumvent this problem.

In our implementations, users can choose if the Kinect needs to detect and treat twenty points of the skeleton (ten in the upper body and ten in the lower portion), or they can choose the seated mode which only detects and treats the ten points of the upper skeleton ignoring the ten points of the bottom.

3.2 Who’s controlling the application?

We use a RGB camera in applications to identify if any skeleton is detected and, if there are different people, we use different colours for each skeleton detected. Only the first one to be detected will control the application, ignoring all the others. When the first person leaves the detection area, the second person caught becomes the first and thus controls the application.

3.3 Improving detection

We decided to create two modes of gesture tracking of the skeleton in the applications developed.

In the first mode we used a direct tracking in which the value of the position of the real world is placed in the position of the cursor. The position is recalculated according to the resolution of the window or application area where you want to use the Kinect. Despite being more fluid this mode is more susceptible to interference from sunlight.

In the other mode the last thirty second positions of the skeleton are saved, and after several tests we decided to treat only the last ten positions (corresponding to the last 1/3 second) and making a calculation of the weighted average to ascertain whether the movement is acceptable or if it is an incorrect reading. If it is acceptable the mean value is placed at the coordinates of the cursor.

3.4 Range of user interaction

• In the applications it is possible to change the size of the gesture. This affects the extent of the hand movement required to move the cursor from one side to the other of the window, scaling between the position of the real world and the window.

• The recommended distance to interact with the Kinect sensor is 1.2m (minimum) and 4m (maximum) [5], so we decided to check in real time the distance between the user and the Kinect sensor. To guaranty a more robust interaction, if the user is not between 1.2 and 2.5m the application warns the user if he/she is too close or too far away from the sensor.

4 Conclusion

This project has shown how it is possible to control any type of application using a set of gestures, a set of existing words in the Windows grammar or both. The same gestures can be used for different interactions, alternating the specific function by a word.

It was also shown that Kinect could be used to improve and facilitate the way people with limited mobility can interact with computer applications.

References


