RECPAD 2015

21st Edition of the

Portuguese Conference on Pattern Recognition

University of the Algarve

Instituto Superior de Engenharia

Edited by

João M. F. Rodrigues

Pedro J. S.Cardoso

Roberto Lam

Mauro Figueiredo

Faro, October 30, 2015

UAlg ISE
UNIVERSIDADE DO ALGARVE
INSTITUTO SUPERIOR DE ENGENHARIA

CVTAL

APRP

RECPAD 2015

21st Edition of the

Portuguese Conference on Pattern Recognition

University of the Algarve

Instituto Superior de Engenharia

Edited by
João M. F. Rodrigues
Pedro J. S. Cardoso
Roberto Lam
Mauro Figueiredo

Faro, October 30, 2015

Conference Program

PROGRAMME OVERVIEW

29TH OF OCTOBER

20h15 - 20h30	Registration*	EVA Hotel restaurant hall
20h30	Welcome dinner*	EVA Hotel restaurant

* - Limited to the registrations with dinner included (Localization: `http://goo.gl/uA3fCw`)

30TH OF OCTOBER

Venue: Instituto Superior de Engenharia
University of the Algarve, Campus da Penha (`http://goo.gl/GhK8iA`)

09:30 - 10:00	Registration	Room 6
10:00 - 10:15	Welcome session	Anf. José Silvestre
10:15 - 11:00	Poster session I	Hall
11:00 - 11:30	Coffee break	Room 6
11:30 - 12:30	Poster session II	Hall
12:30 - 14:00	Lunch	
14:00 - 15:00	Invited Talk	Anf. José Silvestre
Deep Hierarchies in Human and Computer Vision, by prof. Norbert Krüger, University of Southern Denmark		
15:00 - 16:00	Poster session III	Hall ‖ APRP assembly (Anf. José Silvestre)"
16:00 - 16:30	Coffee break	Room 6
16:30 - 17:00	APRP - Thesis Awards, Best Poster Award and Closing Session	Anf. José Silvestre

COMMITTEES

TECHNICAL COMMITTEE

- Luís A. Alexandre (UBI)
- Hélder Araújo (UC)
- Jorge Barbosa (FEUP)
- Aurélio Campilho (FEUP)
- Jaime Cardoso (FEUP)
- Joaquim Pinto da Costa (FCUP)
- Mário Figueiredo (IST)
- Ana Maria Mendonça (FEUP)
- Pedro Pina (IST)
- António Pinheiro (UBI)
- Armando Pinho (UA)
- Bernardete Ribeiro (UC)
- João Barroso (UTAD)
- Paulo Salgado (UTAD)
- João Sanches (IST)
- Luís Silva (UA)
- Paulo Oliveira (IST)
- João Tavares (FEUP)
- Luís F. Teixeira (FEUP)
- Ana Maria Tomé (UA)
- Verónica Vasconcelos (ISEC)
- Andrzej Wichert (IST)
- Fernando Monteiro (IPB)
- Paulo Carvalho (UC)
- Noel Lopes (IPG)
- Catarina Silva (IPL)
- João Paulo Costeira (IST)
- Hans du Buf (UALG)
- Alexandre Bernardino (IST)
- Ana Fred (IST)
- Jorge Santos (ISEP)
- Hugo Proença (UBI)

ORGANIZING COMMITTEE

- João Rodrigues
- Pedro Cardoso
- Roberto Lam
- Mauro Figueiredo

LOCAL TECHNICAL SUPPORT

- Gisela Oliveira

FOREWORD

The 21st edition of the Portuguese Conference on Pattern Recognition, RECPAD 2015, is held at the Universidade of the Algarve (UAlg), Faro, Portugal on the 30th of October, 2015. It is a great honour for UAlg and for the members of the Organizing Committee to have this opportunity to put together this conference.

From the 32 received submissions, 30 papers were accepted. All submissions were double blind and were send to be reviewed by three members of the Technical Committee. All papers had at least one review feedback, most of them 2 or 3 reviews. The conference closing session will include the Best Paper Award and also the ceremony of the APRP Master Thesis Award.

An invited lecture by Prof. Norbert Krüger, Maersk Mc-Kinney Moller Institute for Production Technology, University of Southern Denmark, will present a talk on Deep Hierarchies in Human and Computer Vision.

We are very happy to have the support of the following sponsors: Eva Hotel which helped us in the hotel and dinner conference logistics, and SPIC – Creative Solution for all the layouts and graphics.

On behalf of the organising committee, thank you to all the people involved to this event, namely, the members of the Technical Committee, the Portuguese Association for Pattern Recognition, APRP, specially its president, Prof. Jaime S. Cardoso and to the University of the Algarve – Instituto Superior de Engenharia, with a special thanks to Prof. Ilídio Mestre, director of the institute which will held the conference. Finally, we would like to thanks the CINTAL – *Centro de Investigação Tecnológica do Algarve*, and the precious help from Dr[a]. Gisela Oliveira, with all the work related to the registrations and invoices.

We hope you enjoy this year's edition of RECPAD.

INVITED TALK

TITLE

Deep Hierarchies in Human and Computer Vision

PRESENTER

Norbert Krüger, University of Southern Denmark

CONTACT

norbert@mmmi.sdu.dk

BIOGRAPHY

Professor, Ph.D., Maersk Mc-Kinney Moller Institute for Production Technology, Technical Faculty at the University of Southern Denmark.

Norbert Krüger has been employed at the University of Southern Denmark since 2006 (first as an Associate Professor and then as a full Professor (MSO) since 2008). He is one of the two leaders of the Cognitive and Applied Robotics Group (CARO, caro.sdu.dk) in which currently 12 PhD students, two Assistant and two Associate Professor as well as 8 master students are working. Norbert Krüger's research focuses on Cognitive Vision, in particular vision based manipulation and learning. He has published 45 papers in journals and more than 80 papers at conferences covering the topics computer vision, robotics, neuroscience as well as cognitive systems. His H-index is 24. His group has developed the C++-software CoViS (Cognitive Vision Software) which is now used by a number of groups in national as well as European projects. He is currently involved in 2 European projects as well as 4 Danish projects.

ABSTRACT

Computer vision - although being still a rather young scientific discipline - in the last decades was able to provide some impressive examples of artificial vision systems that outperform humans. However, the human visual system is still superior to any artificial vision system in visual tasks requiring generalization and reasoning (often also called "cognitive vision") such as extraction of visual based affordances or visual tasks in the context of tool use and dexterous manipulation of unknown objects.

Two decades ago, there has been a strong connection between the two communities dealing with human vision research and computer vision. This link however has been somehow lost recently and computer vision has been more and more developed into a sub-field of machine learning. In this talk, I argue that the reason for the superiority of human vision for 'cognitive vision tasks' is connected to the deep hierarchical architecture of the primate's visual system.

The talk is divided into two parts: First, I will give an overview about today's knowledge about the primate's (and by that the human's) visual system primarily based on neurophysiological research. This part is based on the paper (Krüger et al. 2013, IEEE PAMI) and is in particular addressing computer vision and machine learning scientists as audience.

In the second part of the talk, I will describe a three level hierarchical cognitive robot system in which actions are learned by observing humans performing these actions (Krüger et al. 2013, KI). Learning is taking place at the different levels of the hierarchy in rather different representations. On the sensory-motor level, the shape and appearance of objects as well as optimal action trajectories and force torque profiles are represented. On the mid-level, a discrete visual representation based on semantic event chains (Aksoy et al. 2011) bridges towards the planning level, the highest representational level. I will describe different the learning problems on the different representational levels and their interaction.

REFERENCES

E. E. Aksoy, A. Abramov, J. Dörr, N. Kejun, B. Dellen and F. Wörgötter (2011). Learning the semantics of object-action relations by observation. The International Journal of Robotics Research, 30, 1229-1249.

N. Krüger, P. Janssen, S. Kalkan, M. Lappe, A. Leonardis, J. Piater, A. J. Rodriguez-Sanchez and L. Wiskott (2013). Deep Hierarchies in the Primate Visual Cortex: What Can We Learn For Computer Vision? I E E E Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1847-1871.

N. Krüger, A. Ude, H. G. Petersen, B. Nemec, L.-P. Ellekilde, T. R. Savarimuthu, J. A. Rytz, K. Fischer, A. G. Buch, D. Kraft, W. Mustafa, E. E. Aksoy, J. Papon, A. Kramberger and F. Wörgötter (2014). Technologies for the Fast Set-Up of Automated Assembly Processes. KI - Künstliche Intelligenz, 28(4), 305-313.

Temporal Segmentation and Quality Assessment of Digital Colposcopies

Kelwin Fernandes[1]
kafc@inesctec.pt
Jaime S. Cardoso[1]
jaime.cardoso@inesctec.pt
Jessica Fernandes[2]

[1] INESC TEC
Porto, Portugal.
Universidade do Porto
Porto, Portugal.
[2] Universidad Central de Venezuela
Caracas, Venezuela.

Abstract

Cervical cancer remains a significant cause of mortality in low- income countries. Digital colposcopy is a promising and inexpensive technology for the detection of cervical intraepithelial neoplasia. However, diagnostic sensitivity varies widely depending on the doctor expertise. Colposcopies cover four stages: macroscopic view, observation under green light, Hinselmann and Schiller. We focus on the temporal segmentation of the video in these steps and in the ordering of colposcopic images according to their quality. Using a KNN classifier we achieved a precision of 97% in the temporal segmentation. We obtained an accuracy of 76% in the prediction of the relative quality between a pair of frames using SVM.

1 Introduction

Despite the possibility of prevention with regular cytological screening, cervical cancer remains a significant cause of mortality in low-income countries. This being the cause of more than half a million cases per year, and killing more than a quarter of a million in the same period [1].

Digital colposcopy is a promising and inexpensive technology for the detection of cervical intraepithelial neoplasia. The diagnostic sensitivity with these resources ranges from 67 to 98%, depending on the expertise of the doctor [1]. The resection of lesions in the first visit could reduce the costs involved in a scheme of successive visits. Also, it would ensure the appropriate care of patients with poor adherence to treatment.

According to the protocol proposed by the World Health Organization (WHO) [1], detection of preinvasive cervical lesions during a colposcopic screening covers the following steps (see Figure 1): macroscopic view with magnifier white light, followed by observation under green light for diagnosis of aberrant vascularization and then evaluation of the cervical characteristics after exposure to acetic acid solution (Hinselmann) and potassium iodine (Schiller) [1]. Although Hinselmann and macroscopic observation cannot be differentiated on healthy patients, these two steps can be distinguished using contextual information. Throughout the procedure, the expert disturbs the cervix area to achieve better focus, to move from one step to the next, to clean the cervix area, etc. Figure 1 shows four transition frames. These scenes do not bring useful information for the diagnosis and should not be considered in the detection of lesions.

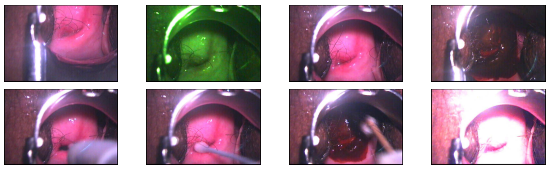


Figure 1: **Top:** Diagnosis steps. From left to right: macroscopic observation, green filter, Hinselmann and Schiller. **Bottom:** Transition frames. The first three frames have occlusions of the cervix area and the last one presents a strong illumination difference after removing the green filter.

In order to decide the right diagnostic, the decision maker (either a physician or an automatic tool) needs to select a subset of good quality frames from each stage. The feedback obtained on Quality Assessment (QA) tasks is usually defined in a numeric scale by subjective ratings [5], which suffers from misinterpretations because of inter-human inconsistencies and limitations of the underlying scale. A more natural way to define quality is induced by a preference function whereby it is defined which object (if any) is the best within an object pair.

The goal of the project is to provide a more effective tool for the diagnosis of pre-invasive lesions, for environments with different resource availability and with different training staff. Our aim is to develop a diagnostic tool that can automatically identify neoplastic tissue from digital images. During the first phase of the study, we aim to achieve automatic recognition of each of the phases mentioned in the colposcopic study of a patient, in order to fine tune the diagnosis of cervical lesions. The temporal segmentation generated by our tool can be used by further techniques to detect lesions.

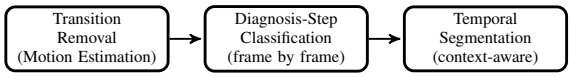


Figure 2: Flow chart describing the proposed framework.

2 Temporal Segmentation

An automated system is proposed in this paper to segment the different steps of the colposcopic assessment. Our system can be splitted in three stages: transition removal, diagnosis-step classification (frame labeling) and temporal segmentation. Figure 2 illustrates this process. The remainder of this section details the proposed system.

2.1 Transition Removal

We assume that transitions correspond to frames with high motion. Therefore, frames with large Euclidean pixel-wise distance with their adjacent frames were filtered [3].

2.2 Diagnosis Step Classification

Then, frames are individually labeled following the protocol stages using a K-Nearest Neighbors model, describing each colposcopic image by their one-dimensional hue and saturation histograms [3]. Similarity between two images is defined by the mean distance -L_1 or Earth Mover's Distance (EMD)- between their histograms. In order to efficiently reduce the presence of noisy objects in the boundaries of the image, we masked the region of interest by removing everything outside a image-centered circle (with diameter equal to 0.75 of the image side).

Given the huge amount of images and the low intra-variance between image within the same video, an equally spaced subset of images was indexed in the KNN knowledge base.

2.3 Temporal Segmentation

Finally, we have to decide the temporal boundaries between the diagnosis steps. For this purpose, let's generalize the problem of temporal segmentation as the problem of recognizing a word (sequence of predicted labels) with minimal accumulated value in a Weighted Finite Automaton (*WFA*).

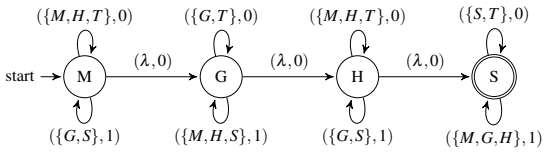


Figure 3: Weighted Finite Automaton that recognizes the temporal segmentation of colposcopies (Transition - T, Macroscopic view - M, Green - G, Hinselmann - H and Schiller - S.

The *WFA* is derived from the domain-dependant protocol. Furthermore, the transition weights are related to the presence of mislabeling. Moreover, if any transition in our policy either consumes an input character or moves "forward" to another state in a directed acyclic graph, the recognition problem holds the conditions to formulate a Dynamic Programming (DP) implementation. Figure 3 shows a graphical representation of the automaton. We denote each phase by its first letter. Since the number of steps in the colposcopic procedure is constant, the DP implementation of the algorithm has a linear performance in the number of frames in the sequence.

2.4 Results

We gathered a dataset of 56 colposcopies from different patients (more than 140k images) annotated by a specialist. Every patient was equally weighted in the compilation of the results. For the assessment of the temporal segmentation we used a leave-one-patient-out cross-validation approach. Transition frames were correctly identified with an accuracy of 90.86% and a precision of 91.46%. Table 1 shows the classification metrics for the temporal segmentation.

Table 1: Average classification metrics per class: Macroscopic, Green, Hinselmann and Schiller. Results with 16 indexed frames per video.

Phase	Distance	Transition				Non-Transition			
		Acc.	Prec.	Rec.	F	Acc.	Prec.	Rec.	F
Macro	L_1	**0.96**	**0.99**	**0.78**	**0.84**	**0.98**	**1.00**	**0.95**	**0.95**
	EMD	0.95	**0.99**	0.74	0.80	0.96	**1.00**	0.89	0.89
Green	L_1	**0.97**	**0.98**	0.66	0.74	0.99	**1.00**	0.96	0.96
	EMD	**0.97**	0.97	0.63	0.70	0.99	0.99	0.91	0.90
Hins	L_1	**0.92**	**0.96**	**0.79**	**0.81**	**0.92**	**0.98**	**0.89**	**0.88**
	EMD	0.91	0.93	0.76	0.77	0.89	0.95	0.86	0.83
Sch	L_1	**0.91**	**0.83**	**0.61**	**0.65**	**0.89**	**0.89**	**0.93**	**0.82**
	EMD	0.89	0.77	0.55	0.55	0.84	0.84	0.83	0.71
Avg.	L_1	**0.94**	**0.94**	**0.71**	**0.76**	**0.95**	**0.97**	**0.93**	**0.90**
	EMD	0.93	0.91	0.67	0.70	0.92	0.94	0.87	0.83

3 Quality Assessment

The quality of a given medical image can be understood from many different perspectives, in our case, the quality is related to how appropriate is the image to diagnose the patient. Usually, this problem has been studied from a subjective numeric or ordinal point of view, wherein an image is shown to the observer, whom is required to give a subjective linearly-ordered rating to the image (e.g. poor, fair, good, excellent; 1 to 5 stars; etc) [5]. A different way to acquire individual opinions is by means of pairwise comparisons, where the observer chooses the best image from a pair. Although the latter approach induces a quadratic number of potential annotations [4, 6], in general, the time needed to judge a given pair is less than the time required to annotate a single image using an ordinal scale [4]. Moreover, the pairwise comparison has been found to be more accurate [4].

However, prediction of image quality from a pairwise perspective has been unexplored. In this sense, we are interested in predicting the quality of a given set of images based on previously annotated images. Training annotations can be obtained in both ways, ordinal and paired, as the ordinal annotations derive a subset of the possible paired annotations. Then, we extract a set of features from the images, and train a predictive model to decide which one is the best image in a given pair.

3.1 Feature Extraction

We extracted a preliminary set of features from the 56 colposcopies that were used in the previous stage of the system. From each colposcopy, we extracted one image per phase. The set of features that were extracted are described below.

- **Image focus:** number of edge pixels in the image.
- **Specular reflection:** size of the largest blob and number of blobs with specular reflections, size of the total area with specular reflections. These regions were detected using the technique proposed by Das et al. [2]
- **Color statistics:** histogram of the hue channel in the HSV color space, average and standard deviation of the three channels in the HSV color space.
- **Transition:** distance to the nearest transition interval before and after the current frame.
- **Cervix:** area occupied by the cervix, center of mass of the cervix (Das et al. [2]) and center of mass of the external orifice of the uterus (i.e. external os).

3.2 Encoding and Model

Finally, we encode the decision problem between a pair of images by the difference between the feature vectors of both images. Thus, the decision problem is transformed from a edge prediction in graph to a traditional binary classification task, which goal is to predict if the first image is better than the second image or not. Thereby, any traditional classification model can be used to solve this problem.

3.3 Results

For the evaluation of the QA we annotated 798 pairs of images and performed experiments using a 10-fold cross-validation. We obtained a preliminary accuracy of 76.2% using a Support Vector Machine with Radial Basis Function (RBF) kernel.

4 Conclusions and Future Work

In this work we provided a framework to temporarily segment a colposcopic assessment. The proposed framework achieved a precision of 91.46% in the transition detection and a precision of 97% in the temporal segmentation. As we observed in the experiments, L_1 distance behaved better than the EMD. Also, we obtained preliminary results in the QA task by means of pairwise ordering of frames. Further work will be done in the quality prediction to force symmetry and transitivity constraints, either in the decision region or by means of some post-processing in the resulting graph of relative quality. Furthermore, we will assess the possibility to predict an ordinal scale of quality from pairwise annotations.

Acknowledgements

This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

[1] Guía global para la prevención y control del cáncer cervicouterino. Technical report, International Federation of Gynecology and Obstetrics, October, 2009.

[2] Abhishek Das, Avijit Kar, and Debasis Bhattacharyya. Elimination of specular reflection and identification of roi: The first step in automated detection of cervical cancer using digital colposcopy. In *Imaging Systems and Techniques (IST), 2011 IEEE International Conference on*, pages 237–241. IEEE, 2011.

[3] Kelwin Fernandes, Jaime S Cardoso, and Jessica Fernandes. Temporal segmentation of digital colposcopies. In *Pattern Recognition and Image Analysis*, pages 262–271. Springer, 2015.

[4] Rafał K Mantiuk, Anna Tomaszewska, and Radosław Mantiuk. Comparison of four subjective methods for image quality assessment. In *Computer Graphics Forum*, volume 31, pages 2478–2491. Wiley Online Library, 2012.

[5] Stefan Winkler. Analysis of public image and video databases for quality assessment. *Selected Topics in Signal Processing, IEEE Journal of*, 6(6):616–625, 2012.

[6] Qianqian Xu, Tingting Jiang, Yuan Yao, Qingming Huang, Bowei Yan, and Weisi Lin. Random partial paired comparison for subjective video quality assessment via hodgerank. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 393–402. ACM, 2011.

The Effect of Symmetry Features on Cerebral Vascular Accident Detection Accuracy

Elmira Hajimani
ehajimani@csi.fct.ualg.pt
University of Algarve, Faro, Portugal

António E. Ruano
aruano@ualg.pt
Centre for Intelligent Systems, IDMEC, IST, Portugal

M. Graça Ruano
mruano@ualg.pt
CISUC, University of Coimbra, Portugal

Abstract

Existence of Cerebral Vascular Accident (CVA) can affect the symmetrical property of human brain, observable on Computer Tomography (CT) images. Analysis of symmetry features is hereby proved to promote the accuracy of classifiers designed for automatic detection of CVA. We attained a reduction up to 40.4% of the number of false detections considering both symmetry features and first and second order statistics, together with model complexity decrease.

1 Introduction

CVA, also called stroke, is caused by the interruption of blood supply to brain, mainly due to a blood vessel blockage (i.e., ischemic) or by hemorrhagic event. The area of the brain that has been affected by ischemic stroke is less dense (darker) than normal area in CT images. In contrast, hemorrhagic stroke is denoted by a denser (lighter) area.

The aim of this work is to investigate the effect of considering proposed symmetry features on the accuracy of classifier designed for automatic detection of CVA within the brain CT images. The result is compared with [1] in which we just used a combination of first and second order statistics to construct the feature space.

Next sections are organized as follows: Section 2 describes how the artifacts are removed from brain CT images. Section 3 discusses the realignment process of tilted images. Symmetry features are proposed in Section 4. Section 5 presents commented experimental results.

2 Removing artifacts

For detecting CVA abnormalities from head CT slices we have to focus on the intracranial part of the images, the part that is inside the skull. Other parts including the scalp, the skull and the U-shaped head holder are considered as artifacts and should be removed. Moreover, those slices which have been taken from the lower part of the head have too much noise from other organs like eyes and nose and contain very small portion of intracranial area. This kind of slices is not also very suitable for CVA detection. The artifact removal process is performed by Algorithm 1, as proposed in [2].

Algorithm 1 Artifact removal algorithm in brain CT images [2]

1. Skull detection:
   - 1.1. Remove pixels whose intensities are less than 250.
   - 1.2. Use Connected Component algorithm [3] to choose the largest component as the candidate skull.
   - 1.3. Remove the small holes within the candidate skull by inverting the matrix of candidate skull and applying the Connected Component algorithm for the second time. Those connected components whose area are less than 200 pixels are considered as hole and will be filled using bone intensity value
2. Removing CT slices with either unclosed skulls or skull containing too many separate regions: Having completed step 1.3, we have already all connected components at hand. As a result we can count the number of big holes (e.g., whose areas are more than 200 pixels). If this number is equal to 2, it will be considered as closed skull otherwise the slice will be removed from the desired set.
3. Intracranial area detection: All CT images that successfully passed step 2 contain only two black regions separated by the skull. To detect which black area is related to intracranial part, the mass centre of the skull is calculated, being considered as intracranial area the region containing the mass centre.

3 Realigning tilted images

Tilted head position in CT images can be either as a part of clinical process (e.g., for reducing beam-hardening artifacts in patients with aneurysm clips[4]) or as a result of patient movement during imaging process. To extract symmetry features we need to detect actual midline of the brain and rotate tilted images to make the actual midsagittal line perpendicular to the x-axis. The method employed is summarized in Algorithm 2 [2, 5]. In order to align tilted CT slices, rotation is done around the mass centre of the skull. Figure 1-b shows the result of applying Algorithms 1 and 2 on the brain CT slice of Figure 1-a.

Algorithm 2 Ideal midline detection of the brain CT [2, 5]

1. Use Algorithm 1 to remove artifacts.
2. Since the concave shape of intracranial region will affect the accuracy of search for finding ideal midline, CT slices with high amount of concavity are found and excluded:
   - 2.1. For each CT slice
     - 2.1.1. Extract the contour of intracranial region.
     - 2.1.2. $Concavity = 0$
     - 2.1.3. For $\emptyset = 0$ to 180
       - 2.1.3.1. Rotate contour by $\emptyset$ degree.
       - 2.1.3.2. $Concavity_{\emptyset} = 0$
       - 2.1.3.3. For $i = 1$ to $number\ of\ rows$
         - 2.1.3.3.1. Scan the pixels of the contour in row i and define the Far Left (FL_i) and Far Right (FR_i) junctions.
         - 2.1.3.3.2. Let C_i be the number of pixels in row i which resides between FL_i and FR_i and is **not** located inside the intracranial region.
         - 2.1.3.3.3. $Concavity_{\emptyset} = Concavity_{\emptyset} + C_i$

       End for
       - 2.1.3.4. $Concavity += Concavity_{\emptyset}$

     End for

     End for
   - 2.2. Sort CT slices based on their corresponding $Concavity$ values and select the first λ slices with least amount of concavity.
3. In order to find the line that maximizes the symmetry of the resulting halves, a rotation angle search around the mass centre of the skull is done:
   - 3.1. For each CT slice remained from step 2:
     - 3.1.1. Let θ be the maximum angle that a given CT image can be tilted.
     - 3.1.2. Let S_j be the symmetry cost at angle j
     - 3.1.3. For $j = -\theta$ to θ
       - 3.1.3.1. Calculate $S_j = \sum_{i=1}^{n} |l_i - r_i|$; where n is the number of rows in the current CT slice, l_i and r_i are the distances between the current approximate midline and the left and right side of the skull edge in row i , respectively.

     End for
     - 3.1.4. Select rotation angle j whose symmetry cost S_j is minimum.

     End for
   - 3.2. The final rotation degree for all CT slices is determined as the median value of rotation angles obtained for each CT slice.

4 Symmetry features

Given the ideal mid-sagittal line, the proposed symmetry features aim on comparing one side of the brain to the other side and discover if

there are any suspicious differences. To extract symmetry features, a window w_1 of size $s \times s$ centred at the pixel (x, y) marked by a clinical expert as normal or abnormal and its contralateral part with respect to the midline, window w_2 centred at the pixel (x', y'), are considered (Please see Figure 1-b). Having identified w_1 and w_2, we can then specify how similar these two regions are by calculating Pearson Correlation Coefficient (PCC) as stated in (1). L_1 norm and squared L_2 norm are also two dissimilarity measures that can be calculated through (2) and (3) respectively. Comparing the intensity value of the pixel that is marked by the expert and its corresponding pixel in the contralateral part can give us another symmetry feature that is stated in (4).

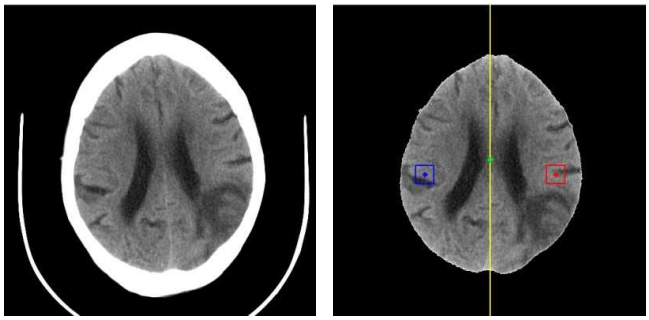
$$PCC = \frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{s}\left(\frac{I_{w_1}^{i,j}-\mu_{w_1}}{\sigma_{w_1}}\right)\left(\frac{I_{w_2}^{i,j}-\mu_{w_2}}{\sigma_{w_2}}\right) \quad (1)$$

$$L_1 = \sum_{i=1}^{s}\sum_{j=1}^{s}|I_{w_1}^{i,j} - I_{w_2}^{i,j}| \quad (2)$$

$$L_2^2 = \sum_{i=1}^{s}\sum_{j=1}^{s}\left(I_{w_1}^{i,j} - I_{w_2}^{i,j}\right)^2 \quad (3)$$

$$diff = I_{w_1}^{x,y} - I_{w_2}^{x',y'} \quad (4)$$

Where $I^{i,j}$ is the intensity value of pixel located at (i, j) within the corresponding window; μ_{w_1}, μ_{w_2}, σ_{w_1} and σ_{w_2} are the mean and standard deviation of intensity values within window w_1 and its contralateral part, window w_2, respectively.



(a) *(b)*

Figure 1: (a) Original brain CT image; (b) After skull removal and realignment, ideal midline is drawn in yellow colour. The green point shows the mass centre (centroid) of skull which the rotation is done around. A window of size 31x31 is considered around pixel located at (365,279) and shown in red colour; its contralateral part with respect to the midline is shown in blue colour.

5 Experimental results

5.1 Dataset

250 CT slices from 14 clinical cases were used [6]. Within these CT slices, 702 pixels were marked as normal or abnormal by Neuroradiologists. (477 normal and 225 abnormal points). In all experiments the size of training, test and validation sets was set to 400, 200 and 102 respectively.

5.2 The effect of using symmetry features

This work uses MOGA[7] to determine the architecture of the Radial Basis Functions Neural Network, its corresponding parameters and input features according to the multiple objectives imposed and their corresponding restrictions and priorities. Table I shows the specification of 2 scenarios. The first scenario is the one which had the best results among the scenarios discussed in [1] using just first and second order statistics as the features (i.e., 23 features in feature space). On second scenario we added up 10 symmetry features by calculating features presented in (1)-(3) for three different window size (i.e., = {11,21,31}) as well as the symmetry feature presented in (4). The number of generations was set to 1000 for both experiments. The number of individuals in each generation was set to 100. In both experiments convex points exist in training set [8]. On first experiment convex hull is applied considering target values but on second experiment the target value is ignored while detecting convex points. In Table 1, CP is the number of convex points in training set; FN_x^{pr} is the number of False Negatives; FP_x^{pr} is the number of False Positives and MC^{pr} states the Model Complexity. TE and TR represent Test and Training sets respectively. The higher value for pr states the higher priority for the corresponding objective.

TABLE I. SCENARIOS OF EXPERIMENTS.

Exp.	Objectives	CP
1	$FP_{TR}^0 \leq 1, FN_{TR}^0 \leq 9, FP_{TE}^0 \leq 16, FN_{TE}^0 \leq 12, MC^0$	329
2	$FP_{TR}^0 \leq 1, FN_{TR}^0 \leq 9, FP_{TE}^0 \leq 15, FN_{TE}^0 \leq 12, MC^0$	356

Table II compares the two experiments with respect to minimum, average and maximum values of FDs (i.e., $FNs + FPs$) and model complexity for best models in non-dominated sets. As we can see, the contribution of symmetry features has promoted the classification accuracy and decreased the average of model complexity. As it can be seen in Table III, the contribution of symmetry features (i.e., experiment 2) reduced the number of false detections within the whole dataset up to 40.4% (from 47 to 28), comparing the best obtained model in non-dominated set. We also had a reduction in model complexity.

TABLE II. MIN, AVG. AND MAX VALUES FOR FALSE DETECTIONS AND MODEL COMPLEXITY OF BEST MODELS IN NON-DOMINATED SET

		TR	TE	V	All	MC
Exp1	Min	0	23	10	47	28
	Avg.	8.60	39.82	21.30	69.73	179.56
	Max	51	53	32	102	300
Exp2	Min	0	11	10	28	105
	Avg.	0.57	22.03	18.30	40.92	152.5
	Max	11	31	24	51	247

TABLE III. MODELS WITH MINIMUM NUMBER OF FALSE DETECTIONS WITHIN WHOLE DATASET

	TR			TE			V			All			MC
	FP	*FN*	*FD*	*FP*	*FN*	*FD*	*FP*	*FN*	*FD*	*FP*	*FN*	*FD*	
Exp1	0	2	2	20	11	31	9	5	14	29	18	47	195
Exp2	0	2	2	4	7	11	8	7	15	12	16	28	154

References

[1] E. Hajimani, M. G. Ruano, and A. E. Ruano, "MOGA design for neural networks based system for automatic diagnosis of Cerebral Vascular Accidents," in *Intelligent Signal Processing (WISP), 2015 IEEE 9th International Symposium on*, 2015, pp. 1-6.

[2] X. Qi, A. Belle, S. Shandilya, W. Chen, C. Cockrell, Y. Tang, *et al.*, "Ideal Midline Detection Using Automated Processing of Brain CT Image," *Open Journal of Medical Imaging,* vol. 3, p. 9, 2013.

[3] L. He, Y. Chao, K. Suzuki, and K. Wu, "Fast connected-component labeling," *Pattern Recognition,* vol. 42, pp. 1977-1987, Sep 2009.

[4] J. H. Brown, E. S. Lustrin, M. H. Lev, C. S. Ogilvy, and J. M. Taveras, "Reduction of aneurysm clip artifacts on CT angiograms: A technical note," *American Journal of Neuroradiology,* vol. 20, pp. 694-696, Apr 1999.

[5] W. Chen, R. Smith, S.-Y. Ji, K. R. Ward, and K. Najarian, "Automated ventricular systems segmentation in brain CT images by combining low-level segmentation and high-level template matching " *BMC Medical Informatics and Decision Making,* vol. 9, p. 14, 2009.

[6] E. Hajimani, C. A. Ruano, M. G. Ruano, and A. E. Ruano, "A software tool for intelligent CVA diagnosis by cerebral computerized tomography," in *Intelligent Signal Processing (WISP), 2013 IEEE 8th International Symposium on*, pp. 103-108.

[7] P. M. Ferreira and A. E. Ruano, "Evolutionary Multiobjective Neural Network Models Identification: Evolving Task-Optimised Models," *New Advances in Intelligent Signal Processing,* vol. 372, pp. 21-53, 2011.

[8] H. R. Khosravani, A. E. Ruano, and P. M. Ferreira, "A simple algorithm for convex hull determination in high dimensions," presented at the 8th International Symposium on Intelligent Signal Processing (WISP), Portugal, 2013.

A SVM Model based on Network Traffic Prediction for Detecting Anomalies

Bruno L. Dalmazo
dalmazo@dei.uc.pt
João P. Vilela
jpvilela@dei.uc.pt
Marilia Curado
marilia@dei.uc.pt

CISUC, Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal

Abstract

Cloud computing is a natural evolution of distributed computing combined with service-oriented architecture. However, its broad adoption has been hampered due to the lack of security mechanisms. Facing this issue, this work aims to propose a new approach for detecting anomalies in the cloud network traffic. The anomaly detection mechanism works on the basis of a Support Vector Machine (SVM) model for binary classification. The key point to improve the accuracy of the SVM model, in the cloud context, is the set of features. In light of this, we present the Poisson Moving Average predictor as the feature extraction approach that is able to cope with the vast amount of information generated over time. We evaluate the performance of our mechanism and compare it against similar studies in the literature, resorting to a real case validation study.

1 Introduction

In the virtual environment, online threats are constantly evolving. Furthermore, cloud computing introduces significant new paths of attack. Distributed Denial of Service (DDoS) is a well-known type of attack that disrupts online operations. Usually, the assault is performed by hundreds (or thousands) of requests for service and it has to be detected before it breaks down the server. Due to the high number of simultaneous requests, this attack generates an anomalous behaviour in the network traffic. Nevertheless, the elastic and scalable nature of the cloud environments is also apt to undergo sudden changes [1], making it even harder to detect which parts of the incoming traffic are from vandalism or legitimate usage.

Several techniques have been already proposed to perform anomaly detection in the cloud environment, such as fuzzy logic, artificial neural network and decision tree classifier. Also, different types of network traffic information are used to detect anomalies, such as protocols behaviour, CPU utilization and user logs. However, there is an apparent deficiency in detecting anomalies with low rates of false alarms. In particular, these techniques require extensive tuning for improving the sensitivity and achieving satisfactory results. There is also no consensus about the best set of features which should be monitored in the network. Moreover, attackers have been able to study the mechanisms behind these techniques and adapt the attack behaviour to evade identification. In this context, the literature lacks mechanisms able to improve the accuracy of the anomaly detection for cloud while keeping a low false detection rates.

To overcome these gaps, a new approach to detect anomalies in a cloud environment is proposed. Our proposal resorts to traffic prediction to generate features that represent the expected proper behaviour of the network traffic. This information is then used jointly with a Support Vector Machine (SVM) model that is fed with the features extracted from network traffic prediction. The combination of these two tools represents a novel and effective approach for detecting anomalous events in the cloud environment. The forecast is performed by a statistical method based on a Poisson process, that has shown itself suitable for dynamic environments such as cloud computing [3]. SVM is already known as one of the best learning algorithms for binary classification [4]. Binary classification meets the goal of this proposal, since we aim to identify the anomalies inside the normal network traffic.

The remainder of the paper is organized as follows. Section 2 covers some of the most prominent related work. Section 3 describes the proposed solution and the methodology used for this paper, whilst Section 4 presents the evaluation and discusses the results. Section 5 concludes with some final remarks and prospective directions for future research.

2 Related Work

This section presents the latest research findings on SVM models applied in intrusion detection system context. Mulay *et al.* [7] presented an IDS that combines SVM and decision trees to build a multi-class SVM. This model can classify the network traffic in normal or abnormal. Horng *et al.* [6] proposed a Network Intrusion Detection System on the basis of Support Vector Machine with features selected by a hierarchical clustering algorithm. The DARPA dataset was used to evaluate the proposed IDS.

Shon and Moon [8] presented a hybrid machine learning approach to detect anomalies in the network traffic. This model is a blending between supervised and unsupervised SVM model. Besides, they use a Genetic Algorithm for extracting more appropriate packet fields (protocol, source port, IP, TTL). Chen *et al.* [2] did a comparative study between Artificial Neural Network (ANN) and Support Vector Machine to predict attacks on the basis of frequency-based encoding techniques to select the features. The results have shown that both approaches are able to detect anomalies in the network traffic, but SVM outperforms ANN.

The key point for using the SVM model with success, in the cloud context, is finding the proper feature extraction approach able to deal with the vast amount of information generated over time. In summary, there is no traditional anomaly detection system to meet these requirements efficiently, since the cloud computing environments have their particular nature and essence. In order to deal with these limitations, in the following section we introduce a conceptual solution for detecting anomalies in the cloud network traffic, by means of a Support Vector Machine fed with features gathered from the Poisson Moving Average predictor.

3 Anomaly Detection Mechanism

The purpose of our *Anomaly Detection Mechanism* is to provide an efficient method to detect anomalies in the cloud-based network traffic. Figure 1 depicts the basis of our mechanism, by highlighting the application scenario and the main conceptual components.

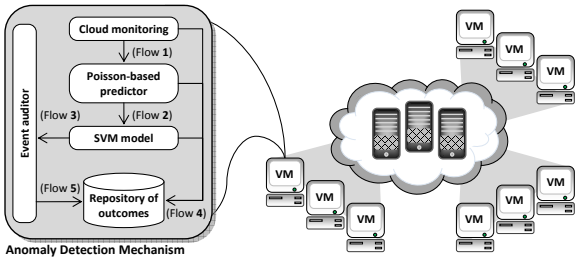


Figure 1: Application scenario and elements of the proposed mechanism

The cloud provider offers several services by the Internet, such as infrastructure, software and platform to the clients. Real-time cloud traffic data (Flow 1) is continuously being gathered from the cloud environment by the *Cloud Monitoring* module. This information is subsequently processed by the *Poisson-based Predictor* that performs prediction based on information such as the protocol type, the number of network packets and timestamp.

After that, the *SVM Model* is fed with features extracted from the predicted data (Flow 2). Then, the *SVM Model* triggers a warning to the *Event Auditor* when an anomalous behaviour is detected (Flow 3). In the meantime, the *Repository of Outcomes* component stores a detailed output regarding the historic of the Virtual Machine (VM) operation (Flow 4). Furthermore, the *Event Auditor* represents an agent placed in the VM

that is able to communicate collaboratively with agents in the other VMs. This agent receives any anomalous event from the *SVM Model* and builds a message with information of all components (Flow 5) for sending alerts to other agents.

Having presented an overview of the anomaly detection mechanism, in the following subsections there will be a more detailed description of the forecasting approach for estimating network traffic on the basis of a Poisson process and the Support Vector Machine model for detecting anomalies in the cloud-based environment.

3.1 Poisson-based Predictor

Predicting the network traffic is an important instrument to support a better understanding of the network traffic behaviour. In light of this, the *Poisson-based Predictor* component represents the feature extraction approach. For, after that, it feeds the SVM model.

The predictor receives as input, the time series created in the *Cloud Monitoring*. Therefore, the predictor is able to forecast the expected value in the network traffic according to the temporal granularity of the time series. We consider prediction based on the Poisson Moving Average (PMA) because it has shown to be suitable for dynamic cloud environments [3]. At this point, the predictor will generate several features, as the outcome of this process. The set of variables used in this approach to describe the network traffic are: the time, the type of protocol, the number of packets, the predicted data and the variance between real network traffic and the predicted network traffic.

3.2 SVM Model

The Support Vector Machine (SVM) model uses a methodology for choosing the best hyperplane (among many others) that represents the largest margin between two classes, namely, normal network traffic and anomalies in this work. Then, the hyperplane is chosen such that the distance from it to the nearest support vector on each side is maximized [4].

The Support Vector Machine learning model includes two stages: training and testing. The first learns the two possible patterns of the network traffic (the normal and the anomalous behaviour). The second tests the knowledge achieved in the past stage to detect unknown anomalies. Separating data into training and testing data is an important part of validating the SVM model. By this, we can minimize the effects of data inconsistencies and better understand the characteristics of the model. Once the SVM model has been processed by using the training set, it is needed to evaluate the prediction capability against the training set. Considering the data in the testing set already contains known values for the attribute that we want to predict, it is possible to determine whether the model's suggestions are correct.

In short, the anomaly detection for the cloud network traffic based on SVM with PMA expresses a process of pattern recognition. In this process, the training data represents the standard pattern and the testing data alludes to identify such pattern. The process of identifying a particular behaviour inside of the testing data is a mapping process of the testing data in some existing pattern of the training data.

4 Evaluation

We consider the DARPA dataset [5] for evaluating this proposal. Table 1 shows the comparison among several approaches that use SVM and DARPA dataset to validate the model. Regarding detection rate (DR) point of view, the best model is proposed by Chen W. *et al.* [2], but this approach showed more than 10% of false positive rate (FPR). Also, the omission of FNR results hampers a better evaluation of this approach's performance.

Another model with high DR, but low FPR, is the Soft margin SVM with Radial Basis Function (RBF) kernel. This model obtained 98.65% of DR, but at cost of high false negative rate (FNR), more than 11%. Other models presented in the Table 1 present at least one drawback: low accuracy, high FPR or high FNR.

In summary, our method on the basis of SVM and RBF kernel with features extracted from Poisson Moving Average predictor presents the best equilibrium in the results. It reaches 98.56% of detection rate and 8% of FNR. Also, our approach displays the lowest FPR among the related work, just 1.44%.

Table 1: Approaches that use SVM and DARPA dataset

Approach	Kernel	DR(%)	FPR(%)	FNR(%)
LIBSVM and PMA	RBF	98.56	1.44	8.00
Horng S. *et al.* [6]	RBF	95.72	N/A	N/A
Soft margin SVM	Inner product	90.13	10.55	4.36
Soft margin SVM	RBF	98.65	2.55	11.09
Soft margin SVM	Sigmoid	95.03	3.90	12.73
One-class SVM	Inner product	53.41	48.00	36.00
One-class SVM	RBF	94.65	20.45	44.00
Enhanced SVM [8]	Sigmoid	87.74	10.20	27.27
Chen W. *et al.* [2]	RBF	100.00	10.35	N/A

5 Conclusions and Future Work

In this paper, we have shed light on the major problem for preventing the adoption of the cloud service models: security (or lack thereof). In this context, a novel approach to detect anomalies in the cloud scenario was proposed. The anomaly detection approach relies on a distributed and collaborative mechanism that combines a Support Vector Machine model with features extracted from a Poisson Moving Average predictor.

By analysing the evaluation results, it can be seen that the anomaly detection mechanism was able to identify anomalies considering a case study with real data. Our SVM model achieved high degree of accuracy providing the best compromise in terms of detection and false alarm rates. In particular, our approach exhibits the lowest level of false positive rate and the second best false negative rate in comparison with other approaches. Prospective directions for future research also include: (i) proposing a feature selection approach using Genetic Algorithms; and (ii) proposing an unsupervised Support Vector Machine model.

References

[1] Hitesh Ballani, Paolo Costa, Thomas Karagiannis, and Antony IT Rowstron. Towards predictable datacenter networks. In *SIGCOMM*, volume 11, pages 242–253, 2011.

[2] Wun-Hwa Chen, Sheng-Hsun Hsu, and Hwang-Pin Shen. Application of SVM and ANN for intrusion detection. *Computers & Operations Research*, 32(10):2617–2634, 2005. ISSN 0305-0548.

[3] Bruno L. Dalmazo, Joao P. Vilela, and Marilia Curado. Online traffic prediction in the cloud: A dynamic window approach. In *The 2nd International Conference on Future Internet of Things and Cloud (FiCloud'2014)*, pages 9–14, Aug 2014. doi: 10.1109/FiCloud.2014.12.

[4] Naiyang Deng, Yingjie Tian, and Chunhua Zhang. *Support vector machines: optimization based theory, algorithms, and extensions*. CRC Press, 2012.

[5] J.W. Haines, L.M. Rossey, R.P. Lippmann, and R.K. Cunningham. Extending the darpa off-line intrusion detection evaluations. In *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX '01. Proceedings*, volume 1, pages 35–45 vol.1, 2001.

[6] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, and Citra Dwi Perkasa. A novel intrusion detection system based on hierarchical clustering and support vector machines. *Expert Systems with Applications*, 38(1):306 – 313, 2011. ISSN 0957-4174.

[7] Snehal A Mulay, PR Devale, and GV Garje. Intrusion detection system using support vector machine and decision tree. *International Journal of Computer Applications*, 3(3):40–43, 2010.

[8] Taeshik Shon and Jongsub Moon. A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18): 3799 – 3821, 2007. ISSN 0020-0255.

Comparison of Modelling and Classification Techniques Applied to Marble Classification

João M. Pedro Batista (joao.batista@tecnico.ulisboa.pt)
João R. Caldas Pinto (joao.c.pinto@tecnico.ulisboa.pt)
Susana Vieira (susana.vieira@tecnico.ulisboa.pt)
Pedro M. Amaral (pedro.amaral@tecnico.ulisboa.pt)

IDMEC/LAETA
Av. Rovisco Pais
1049-001 Lisboa

Abstract

Automatic marble classification based on visual appearance is an important and complex industrial issue. However, there is no definitive solution, mainly due to the presence of high number of randomly distributed colors and the subjective evaluation made by human experts. In this work, one of the tasks is to analyze a set of soft computing classification algorithms for a specific database in representation of general types of marbles. The main objective of this research are color and homogeneity classification through binary models algorithms with fuzzy and logistic regression approaches. The features are acquired by *QuadTree* segmentation. The results of this study are very encouraging with fuzzy based on Takagi-Sugeno inference system with fuzzy c-means clustering giving the best score for color model with the minimum of 25 training samples. For homogeneity classification the minimum number of training samples were settled to 20 and the best score obtained was for fuzzy based on Takagi-Sugeno inference system with Gustafson-Kessel clustering.

1 Introduction

Visual classification is a big issue at the last stage of the marble industry process. It is a time consuming process with high subjectivity which was the trigger to a deeper research and development of automatic methods, however without satisfying results so far. The main handicap of this kind of implementation is the high diversity of types of marbles in different geographic regions with a high variety of colors and homogeneities typical of natural stones [1]. The database was obtained by an automatic image acquisition machine at FrontWave Company with stones provided by Grupo Galrão between February 2015 and June 2015. A set of models was obtained from this database, expecting that these can be extended to other manufacturers.

2 Database and Feature Extraction

The samples were organized by experts for each classification criteria, color and homogeneity. Grouping by similarity of the predominant color, homogeneity and background color (typically the lightest color) is the main goal of human classification. As a result of these pre classifications, six different types of color are considered, with an amount of 70, 60, 87, 150, 60 and 60 samples (Figure 1). For homogeneities, three different types were defined with an amount of 98, 287 and 102 (Figure 2).

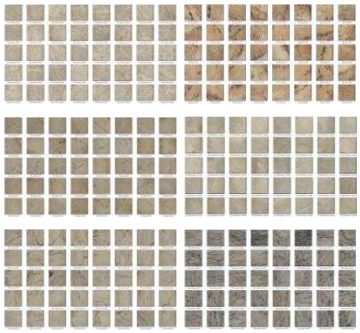


Figure 1: Preview of Marble Color Pre Classification

Figure 2: Preview of Marble Homogeneity Pre Classification

Several modern techniques of color and texture features were tested [2, 3, 4], however there are no considerable improvements on the results when compared with the more classical technique based on a *QuadTree* segmentation [5] of each image. This process was initialized with four segments (*firstSeg*), and the minimum area allowed for each final region (*minArea)*, was 4 pixels. The criterion to parcel out each segment was based on standard deviation of brightness distribution (eq.1).

$$f_k = \sigma_V \geq stdHomog \quad , k = considered\ region \qquad (1)$$

The value of *stdHomog* was varied by 3, 5 and 7. From the segmentation it was possible to extract the four main regions that together define the marble predominant color (green plot). The veins are represented in red plot (Figure 3b), similarly with previous work [6]. The veins are also characterized by its degree of whiteness, represented by artificial colorification (Figure 3c).

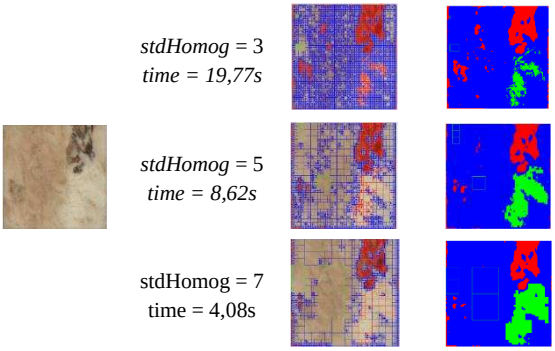


Figure 3: Marble a) Original, b) Segmentation, c) Artificial Color

For color classification the features extracted from this method were the mean and standard deviation of RGB and HSV [7] for both the global image and the total of the predominant regions, totalizing 24 features. Another considered feature was the pixel value corresponding to the maximum gray level histogram of the marble image. For homogeneity classification the features gathered were: Homogeneity (eq.2), the difference between the light and dark veins weighted by the percentage of each of the regions selected, *weiVarVeins* (eq.3) and entropy E (eq.4), totalizing 3 features.

$$Homogeneity = 1 - \frac{minArea.(\sum(regions) - firstSeg)}{Total\ Image\ Area} \qquad (2)$$

$$weiVarVeins = varVligth.\frac{pVeinLight}{pVeins} + varVdark.\frac{pVeinDark}{pVeins} \qquad (3)$$

Where *pVeins* is the percentage of all veins in the image (regions with red plot in Figure 3b). The *pVeinLight* is the percentage of green area and *pVeinDark* is the percentage of red area in the image (Figure 3c). The absolute difference between the mean values from component V of HSV in all the light veins and all predominant regions is *varVligth*. The same is considered for *varVdark* with all dark veins.

$$E = -\sum\left(p.log2(p)\right) \qquad (4)$$

p is the normalized values of gray level histogram of the image (eq.5).

$$p = \frac{h}{\sum(total\ image\ pixels)} \quad , h = hist(image) \qquad (5)$$

3 Classification Methods

All the data obtained in the feature extraction process has to be normalized in order to have features with values between 0 and 1. To reach this goal the standard deviation normalization algorithm [8] was used. The data was divided in training and validation sets using a 5 fold cross validation algorithm [8]. Logistic Regression (LR), the classic linear method for binary classification [8] was tested, as well as intelligent techniques based on Takagi-Sugeno fuzzy models (TS-FM) [9], where each rule describes a local input-output relation. During this work the antecedent fuzzy set was determined by Fuzzy C-Means (FCM) clustering and the Gustafson Kessel (GK) clustering [10, 11] algorithms.

4 Approach

To test all the algorithms is statistical terms, a cross validation algorithm divides the data by 20%, 40%, 60% and 80% for the training set. With 20 runs of each model, the mean and standard deviation of accuracy (ACC), sensitivity (TPR), specificity (SPC) and area under the operation receiver curve (AUC) [12] can be computed to evaluate the algorithms.

The objective of this approach was to conclude which intelligent model performs better for the small number as possible of training data in the lowest computational time (that decreases with higher values of *stdHomog*). The high variability of models performances forced to consider a weighted mean value allowing to infer on the quality of positive and negative samples correctly classified. A higher rank of TPR was assumed due to the industrial context of this work.

$$Weigthed\ Mean\ Score = 0{,}6TPR + 0{,}4SPC \qquad (6)$$

The best percentage of training set is first obtained and then the best model is evaluated for each classification category. The final parameter specified is the *stdHomog* for *QuadTree* segmentation.

5 Results

Given the three different raw data sets, one for each *stdHomog*, the results trend is the same even when varying the training set percentage for all models. The results of the mean and standard deviation, represented by an error bar (Figure 4) were calculated for six models in terms of color classification and three models for homogeneity. All the results for each *stdHomog* are identical.

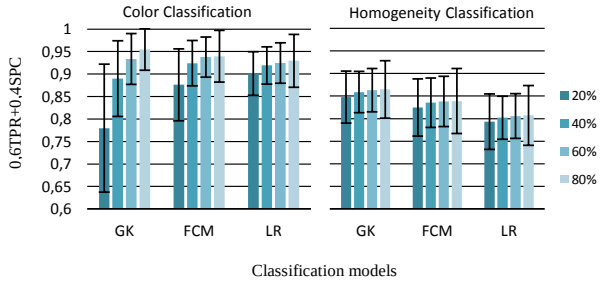


Figure 4: Results of models score with different training percentage for *stdHomog* = 5

In color classification it can be noted an improvement from 20% to 40% and from this percentage on there is no significant model score increase. For the case of homogeneity classification the results have a lower score because the database was unbalanced for different homogeneities by the experts and there is no score improvement with the increasing percentage of training set. In conclusion the best training percentage for color classification is 40% and for homogeneity is 20%.

Figure 5 presents the results of the three different *stdHomog* considered for the training percentage.

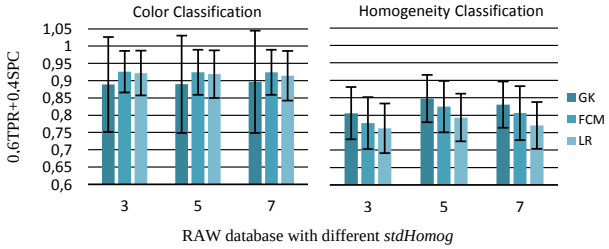


Figure 5: Scores of all database for the different models tested

These results are conclusive for the *stdHomog* to consider in the *QuadTree* image segmentation. For color classification there is no relevant variation to decide which *stdHomog* to choose but for homogeneity classification it is possible to observe an increasing score of 3 to 5 and decrease for 7. To choose the best database for both classifiers, *stdHomog*=5 is the best choice.

The best model for color classification is the FCM TS-FM and for homogeneity classification is the GK TS-FM.

6 Conclusions

This paper compares three different classification models for color and homogeneity of marbles. Different settings for the *QuadTree* algorithm were tested to determine how many samples are needed. The color classification should use at least 25 samples per class for training the FCM TS-FM to have more than 92% of performance. For homogeneity classification using 20 samples to train a GK TS-FM allows to have more than 84% of model performance. The computational time for classification is practically the time that takes to extract all the features for each sample. For a high heterogeneous marble this time is about 12s. Future work will deal with a more deep study of other features for each classification model and test the models with new samples.

Acknowledgements

This work was funded by IST - project 2161, LAETA - UID/EMS/50022/2013 under the scope of the project QREN "*StoneScan*" (2013/30369 "Sistema de Incentivos à Investigação e Desenvolvimento Tecnológico"). This work was supported by FCT, through IDMEC, under LAETA, Program Investigator FCT (IF/00833/2014), co-funded by the ESF through the POPH. Authors want also to thank Engs. Adriano Coelho and Nuno Reis from FrontWave Company.

References

[1] E. M.Winkler, "Weathering and weathering rates of natural stone," *Environmental Geology and Water Sciences, Volume 9, Issue 2,* pp. 85-92, 1987.

[2] T. Alphonsa and K. Sreekumar, "A Survey on Image Feature Descriptors-Color, Shape and Texture," *International Journal of Computer Science and Information Technologies,* vol. V, pp. 7847-7850, 2014.

[3] R. M. Kumar and K. Sreekumar, "A Survey on Image Feature Descriptors," *International Journal Science and Information Technologies,* vol. V, pp. 7668-7673, 2014.

[4] F. Xu and T.-J. Zhang, "Evaluation and comparison of texture descriptors proposed in MPEG-7," *Journal of Visual Communication and Image Representation,* vol. 17, pp. 701-716, 2006.

[5] F. F. De Natale & Granelli, "Structured-based image retriveval using a structured color descriptor," *Proceeding of the International Work-shop on Content-Based Multimedia Indexing,* vol. CBMI'01, pp. 109-115, 2001.

[6] S. M. Vieira, J. M. Sousa and J. R. Caldas Pinto, "Ant based Fuzzy Modeling applied to marble classification," *Image Analysis and Recognition, Third International Conference, ICIAR,* vol. II, pp. 90-101, 2006.

[7] J. Foley, A. Dam, S. Feiner and J. Hughes, Computer Graphics: Principles and Practice, New York: Addison Wesley, 1990.

[8] T. Hastie, R. Tibshirami and J. Friedman, "The Elements of Statistical Learning," in *Data Mining, Inference and Prediction*, Stanford, California, Springer, 2008.

[9] M. Sugeno and T. Yasukawa, "A Fuzzy-Logic-Based Approach to Qualitative Modeling," *Transaction on Fuzzy Systems,* Vols. I,no. 1, pp. 7-31, 1993.

[10] J. C. Bezdek, R. Ehrlich and W. Full, "FCM: The Fuzzy c-Means Clustering Algorithm," *Computer and Geosciences,* Vols. X, no. 2-3, pp. 191-203, 1984.

[11] R. Babusca, P. J. van der Veen and U. Kaymak, "Improved Covariance Estimation for Gustafson-Kessel Clustering," *IEEE International Conference on Fuzzy Systems,* pp. 1081-1085, 2002.

[12] J. A. Hanley and B. J. McNeil, "The Meanind and use of the Area under a Receiver Operation Characteristic (ROC) Curve," *RADIOLOGY,* vol. 143, pp. 29-36, 1982.

Automatic Classification of Heart Sound Signals using a Hidden Markov Model

Jorge Oliveira
oliveira_jorge@dcc.fc.up.pt
Department of Computer Science,
Faculty of Science, Porto University

Miguel Coimbra
mcoimbra@dcc.fc.up.pt
Department of Computer Science,
Faculty of Science, Porto University

Abstract

Hidden Markov Models have been successfully applied in heart sound segmentation and classification. In this paper an ergodic, homogeneous and stationary hidden Markov model is implemented in a public dataset. This model uses a multiple mixture of multivariate Gaussian distribution for the probability state-dependent distribution, the perpetual and spectral features are used as an input to the system.

1 Introduction

The phonocardiogram (PCG) signal is obtained during an auscultation using a traditional or an electronic stethoscope. The PCG contains important information concerning the mechanical activity of the heart valves. The signal processing of a PCG has two main goals; The first one is to split the PCG into the heart cycles, where each heart cycle is composed by the first heart sound (S1), the systolic period, the second heart sound (S2), and the diastolic period (Figure 1). The second goal is the detection of other sounds such as the third and fourth heart sounds (S3 and S4) as well as heart murmurs that may be associated to cardiac pathologies. For heart sound segmentation, the methods can be divided depending on which domain they are applied: the time domain (Shannon energy [6]), frequency domain (homomorphic filter [8]) and entropy domain (entropy gradient [9]). For heart sound classification different classifiers have been proposed Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Hidden Markov Models (HMMs). The HMMs are used to statistical model the highly dynamic and non-stationary nature of the cardiac system. In Gill [13], the signal is pre-processed and a subset of candidates (peaks) are extracted from a homomorphic envelogram, these candidates are classified using a discrete-time HMM, where the state-distribution is modelled using the time-duration from the preceding candidate. Schmidt [14] implemented a duration-dependent HMM using a homomorphic envelogram, which has the advantage (compare to the traditional HMMs) that every state duration is explicitly model in the state transition matrix. Chung [12] detected and classified heart sounds using first a left-right HMM (the first state is assumed to be known) and later an ergodic hidden Markov model (an initial state probability distribution), the spectral variability in each state is modelled using multiple mixture of multivariate Gaussian distribution. In this paper, we implement Chung approach using spectral and perceptual features as an input to the system. The results are based on real examples from a public dataset (Pascal challenge [7]).

2 Hidden Markov Models

Hidden Markov Models are models in which the distribution that generates an observation depends on the state of the underlying and unobserved Markov Process. The Chung ergodic hidden Markov model assumes that the state transition probability matrix Γ is homogenous:

$$\gamma_{ij}(t) = \Pr(C_{s+t} = j | C_s = i), \quad (1)$$

The probability transition do not depend on s. In another words, the probability of being state j after t-steps from state i is independent from the current evaluation time s. This is by itself an unrealistic constrain imposed to the system, simply because the PCG is not a stationary signal and therefore, the statistical state properties are constantly changing over the time. The spectral and perceptual variability at each step is modelled using a multiple mixture of multivariate Gaussian distribution. Let $\delta_1 \dots \dots \delta_m$ denote the probabilities assigned to the different components and let $p_1 \dots \dots p_m$ denote their probabilities. Giving $X(t) \in \mathcal{R}^l$ (l - features), the output probability distribution in state j is given by:

$$P_j(X(t)) = \sum_{i=1}^{m} \delta_{ij}\, p_{ij}(X(t)), \quad (2)$$

The state dependent distribution $p_{ij}(X(t)) = N(X(t); \mu_{ij}, \Sigma_{ij})$ is a multivariate Gaussian distribution with mean vector μ_{ij} and covariance Σ_{ij}. We denote $u(1)$, the initial distribution of the hidden Markov model. It is assumed that each state of HMM correspond to an element of the heart sound signal because the signal characteristics in each element are thought to be homogenous, as it is depicted in Figure 1.

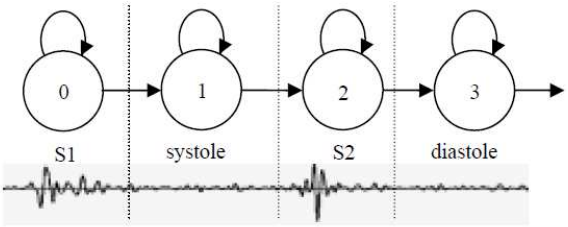


Figure 1: Four state hidden Markov model for a cycle of a normal heart sound signal [2].

This model ignores S3, S4 and murmur sounds, since these are not usually generated in normal conditions. Additionally these sounds are difficult to hear and to record, they are most likely not noticeable in our heart sound data. Our algorithm must be able to estimate the state-sequence $C_1^T = \{c_1 = S1, \dots., c_l = S2, \dots, c_T = Diastoly\}$, which is most likely to produce a given set of observations $X_1^T = \{x_1, \dots., x_T\}$ of length T.

$$S^* = \operatorname{argmax} \Pr(C_1^T | X_1^T, \theta) \quad (3)$$

Where $\Theta = \{\boldsymbol{u(1)}, \boldsymbol{\Gamma}, \mathbf{P}\}$ denotes the model parameters such as the initial state distribution $\boldsymbol{u(1)}$, state transition probability $\boldsymbol{\Gamma}$ and state dependent distribution $\mathbf{P}$. Before trying to solve the equation (3), it is necessary a good parameter initialization. The mixture of a multivariate Gaussian distribution parameters $\{\boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ are estimated using the expectation maximization (EM) algorithm, which assigns posterior probabilities to each component density with respect to each observation [5]. Finding the maximum likelihood estimation for a multivariate Gaussian mixture, is indeed a really difficult task or sometimes impossible if global optimum does not exist. If the dataset is relatively small, the estimated covariance matrix might become singular, or even if a very few number of points are assigned to a specific class due to wrong parameter initialization. In such situations covariance shrinks to a delta function, the log likelihood is infinite and convergence stop leading to ill-conditioned estimation. In our algorithm, this is avoided by constraining the covariance matrices to be positive definite. Finally the Baum-Welch algorithm is implemented in order to re-estimate the parameters of HMMs.

3 Methodology

In this study 90 heart sound segments are used corresponding to a total 1421 annotation examples of S1 and S2, within a time range between 1.2 and 14.7 seconds, the signal is recorded in 16 bits resolution and sampled at 4000 samples per second. According to [1], the frequency spectrum of S1contains peaks in the 10 to 50 Hz range and the 50 to 140

Hz range, while the frequency spectrum of S2 contains peaks in the 10 to 80 Hz range, the 80 to 200 Hz range, and the 220 to 400 Hz range. As a result, this study limits the spectral feature extraction between the frequencies of 10 Hz and 430 Hz. The system filters the original signal using zero-phase Butterworth bandpass filter order 10, with a lower cutoff frequency of 30 Hz and a higher cutoff frequency of 430 Hz. The signal is normalized (mean subtraction) and properly scaled. The signal is then divided into fragments of 0.25-second length and 0.10-seconds of overlap. Since the average of S1 sound is 0.16 seconds, each fragment is spitted into 0.15-seconds frames and shifted by 0.020 seconds. From each frame, 12 filterbank energies are extracted, these represent different energy ranges equally spaced around the mel frequency domain. The average of these frames is used as features for the classifier. The FFT is computed for each fragment and the total spectral energy around [0-140 Hz] is used a feature. Finally, the total temporal energy of the fragment is also used as a feature. The records were divided into a training set (N=45) and a test set (N =45). Features in the training set are extracted by using a 0.20-second centered window around the annotation event. The systolic and diastolic events are extrapolated using the annotated position of S1 and S2 in the heart sound.

4 Results

In this paper, the heart sound signal is discretized in segments of 0.025-seconds length. This window might be long if it is applied to short signals, resulting in a diminished number of samples. As a result, the algorithm might never convergence to the desired solution. Although using a short window degrades the spectral resolution, which is necessary for a good perceptual and spectral feature estimation. The algorithm performance evaluation cannot be done in the regular way of calculating the error percentage on a given test samples set. For each audio sample, the *S1*, *S2*, *Systole* and *Diastole* events are counted and compared with the corresponding annotation file. The expected sequence *S1-Systole S2-Diastole* in each heart cycle is always observed. The evaluation results are given in Table I.

	ANNOTATED	PREDICTED	(%) WC	(%) RC
S1	717	692	0.04	0.93
SYSTOLE	702	670	0.03	0.91
S2	704	641	0.02	0.86
DIASTOLE	635	623	0.07	0.94

Table I: The classification results using 10 Gaussian mixtures.

The results are achieved without using any auxiliary signal, such as the electrocardiogram (ECG) and the data analysis and processing were conducting using Matlab.
The wrong counting (WC) is defined as a ratio of wrong detected event by the number of annotated event:

$$Wc = {}^{Nº Detected\ Event - Nº Annotated\ Event}/_{Nº Annotated\ Event} \quad (4)$$

The right counting (RC) is defined as a ratio of right detected event by the number of annotated event:

$$Rc = {}^{Nº Detected\ Event}/_{Nº Annotated\ Event} \quad (5)$$

The results in Table I reveals that the model is more sensitive to S1 and diastolic than S2 and systolic events. In Figure 2, we show the matching between the state sequences from the Viterbi decoding in the ergodic HMM and the corresponding annotation sequences for a continuous heart sound signal in the normal conditions. The algorithm is perfectly capable of detecting the right sequence of events, although it did not estimated with very accuracy, the state duration in each event, which might be a consequence of using static state transition matrix $\boldsymbol{\Gamma}$. One possible alternative way of modelling time trend and seasonality in the HMMs is to drop the assumption that the Markov chain is homogenous, and assume instead that the state transition matrix $\boldsymbol{\Gamma}$ is dependent on time.

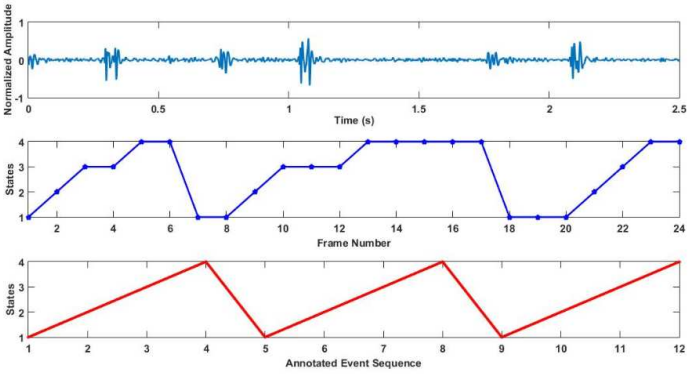


Figure 2: The matching between the state sequence and the annotation sequence for a heart sound signal in normal conditions.

5 Conclusions

In this paper, a heart sound classification algorithm is proposed, which uses an ergodic, homogenous HMM. The spectral and perceptual features were used as input to the system. These features are very descriptive and sensitive to S1 and S2 events, for future work, different feature families (ex: fractal) are going to be added and the optimal feature subgroup determinate. The proposed algorithm seems to be also very suitable for the heart rate detection. A non-homogenous HMM version of the proposed algorithm can strengthen the algorithm in noise presence and enlarge a big variety of problems, such as the detection of S3 and S4, murmurs and valves defects.

References

[1] A.Iwata, N.Ishii, N.Suzumura, K.Ikegaya. Algorithm for detecting the first and the second heart sounds by spectral tracking. Med Biol Eng Comput. 1980 Jan.

[2] J.Yong Chung. Classification of Continuous Heart Sound Signals Using the Ergodic Hidden Markov Model, Proceedings of the 3rd Iberian conference on Pattern Recognition and Image Analysis, (IbPRIA '07), 2007.

[3] D.Gill, N.Gavrieli, N.Intrator. Detection and identification of heart sounds using homomorphic envelogram and self-organizing probabilistic model, Computers in Cardiology, 2005, Sept. 2005.

[4] S.E.Schmidt, E.Toft, C.Holst-Hansen, C.Graff, J.J.Struijk. Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent Hidden-Markov Model. Computers in Cardiology, 2008, Sept. 2008.

[5] G.McLachlanand, D.Peel, Finite Mixture Models. Hoboken, NJ: John Wiley & Sons, Inc., 2000.

[6] H.Liang, S.Lukkarinen, I.Haritmo, Heart Sound Segmentation Based on Heart Sound Envelogram, in Computers in Cardiology, vo.24, 1997.

[7] P.Bentley, G.Nordehn, M.Coimbra, S.Mannor. The PASCAL Classifying Heart Sounds Challenge 2011 (CHSC2011) Results. http://www.peterjbentley.com/heartchallenge/.

[8] C.N. Gupta, R. Palaniappan, S.Swaminathan, S.M.Krishnan,. Neural network classification of homomorphic segmented heart sounds, Applied Soft Computing, 2007.

[9] J. Oliveira, A. Castro, M. Coimbra, "Exploring Embedding Matrices and the Entropy Gradient for the Segmentation of Heart Sounds in Real Noisy Environments", IEEE EMBC, Chicago, USA, 08/2014

Noninvasive Detection of Pulmonary Hypertension by Second Heart Sound Shape Analysis

Ana Castro
ana.castro@dcc.fc.up.pt

Sandra S. Mattos
ssmattos@cardiol.br

Miguel T. Coimbra
mcoimbra@dcc.fc.up.pt

Instituto de Telecomunicações
Faculdade de Ciências da Universidade do Porto
Portugal

Unidade de Cardiologia e Medicina Fetal
Real Hospital Português
Recife, Brasil

Instituto de Telecomunicações
Faculdade de Ciências da Universidade do Porto
Portugal

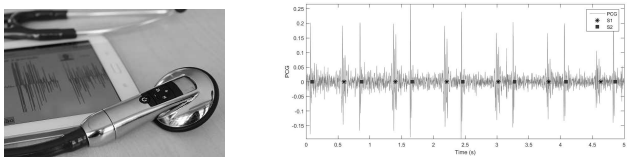
Abstract

Heart sound characteristics are linked to blood pressure, and its interpretation is important for detection of cardiovascular disease. In this study, heart sounds acquired from adult patients with (PH, n=11) and without (NPH, n=32) pulmonary hypertension, were analyzed. A novel method is proposed for the second heart sound analysis, and detection of PH. The proposed method automatically segments the phonocardiogram (first and second heart sounds - S1 and S2), aligning the detected S2s to define the signal envelope for peak-picking and aortic (A2) and pulmonary (P2) components detection (maximum, minimum and difference envelopes). Following, A2 and P2 amplitudes relation was analyzed (P2/A2). The best results in separating NPH from PH patients was obtained using the maximum envelope: NPH: 0.26±0.12, PH: 0.42±0.10 (Wilcoxon Mann-Whitney test, $p<0.01$). The P2/A2 ratio allows to separate the patients with and without PH, and it may be observed that in the NPH group the P2 amplitude rarely surpasses 50% of the A2 amplitude. Larger clinical studies need to be conducted to evaluate the reproducibility of our results. But if proved reliable, this simple algorithm has the potential to allow for a cost-effective tool to detect PH by S2 shape analysis.

1 Introduction

Noninvasive cardiovascular monitoring is gaining a lot of interest for a safe, continuous, and comfortable patient assessment, both in clinical and home-based scenarios [9]. Stethoscopes are part of the first line of screening and diagnosis of cardiac pathologies, nonetheless auscultation is one of the hardest skills to master [4]. Technological developments, such as digital stethoscopes and signal processing, may improve this tool [6]. The typical phonocardiogram (PCG) of healthy subjects is characterized by two main components: the first (S1) and second (S2) heart sounds. S1 corresponding mainly to the closure of the atrioventricular valves, it is low in pitch and relatively long-lasting. S2 occurs when the aortic and pulmonary valves close at the end of the systole (A2 and P2 components); it is characterized by a rapid snap, corresponding to the valves quick closure, and a short period of surrounding vibrations. In pathological cases heart sound may exhibit further components, such as a third (S3) or a fourth sound (S4), or change their characteristics, as for example the hyperphonesis of P2 in Pulmonary Arterial Hypertension (PAH) patients [2, 5]. Pulmonary Arterial Pressure (PAP) may be estimated noninvasively with an echocardiogram, but this estimation is not reliable and, in many cases, difficult to obtain [7]. For precise measurements of PAP, patients are submitted to a right heart catheterization, an invasive procedure with associated discomfort, risks and costs. Noninvasive assessment of PAP through heart sound analysis has been addressed with encouraging results [5, 8], however PAP error estimations are still high, despite elevated computational costs.

Still, the ability to screen for early signs of PAH could direct suspected individuals for further investigations with all the potential benefits from early diagnosis and initiation of therapy. For this purpose, we have been working in the development of tools for digital auscultation [3], with low computational cost algorithms to be implemented in smart-phones or tablets for screening of PAH through PCG analysis. This work presents preliminary results on a new approach for the analysis of the shape of S2, and evaluate the P2/A2 relative amplitude in patients with (PAH) and without (NPAH) PAH.



(a) (b)

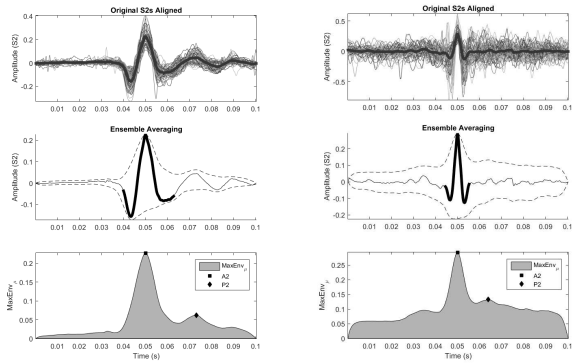
Figure 1: (a) DigiScope Collector system front view. (b) Automatic segmentation of the heart sound (Phonocardiogram, PCG) into first (S1) and second (S2) heart sounds.

2 Dataset

Data were collected at Real Hospital Português (RHP) in Recife, Brazil, with a Littmann® 3200 electronic stethoscope, and the DigiScope Collector software presented in Figure 1(a) (4 kHz, 8 bits resolution) [3]. Data was anonymized and analyzed in Portugal with the approval of the RHP and the University of Porto Ethics Committees. Collected data includes 11 adult patients with PAH and 32 without PAH (NPAH group). Auscultation was performed over the second left intercostal space for a period of 60 s, in a calm and quiet environment. Data analysis performed in MATLAB R2015a. Data presented as mean±standard-deviation.

3 Heart Sound Analysis

The pre-processing of the PCG includes filtering with a band-pass, zero-phase, Butterworth filter order 6 (25-900 Hz) to eliminate out of band noise, decimation to 2 kHz, and normalization. PCGs were automatically segmented, retrieving S1 and S2 markers (Figure 1(b)) [1]. Following heart sound segmentation (77.3±9.3 detected S2s per patient, NPAH: 78.1±9.2, PAH: 74.1±9.7), detected S2s for each patient were aligned using the maximum peak for reference to obtain a S2 average template. The obtained template was then used to align all S2s based on the maximization of the cross-correlation between segments. All aligned segments with a global average correlation inferior to 0.7 were rejected from the analysis, removing possible erroneously marked S2s, or highly contaminated by noise (dissimilar wave morphology). After aligning the S2s in this iterative process (58.5±17.5 S2s per patient, NPAH: 59.3±18.6, PAH: 56.3±14.4), the ensemble average of the segments was obtained. Due to the cyclic nature of the heart sound, and the static patient position during auscultation, one may assume a regular heart rate meaning that each event will be repeatedly represented. By exploiting this assumption, more robust S2 estimations may be obtained, which is especially useful considering the harsh conditions of a real clinical environment (e.g. movement artifacts, respiration sounds, ambient noise). Ensemble averaging is based on a simple signal model in which the observed signal is assumed to be composed of a deterministic component (in this case the S2) and noise. If we assume that the noise is a zero mean stationary process, the ensemble average of the observations may be used for a robust estimation of S2. However, and since S2 varies with respiration, ensemble averaging assumptions may not apply, especially in the pathological cases in which the increased split between A2 and P2 components may lead to signal distortion, and a poor representation of the patient S2 general model. To overcome this issue, positive and negative envelopes (cubic Hermite



(a) (b)

Figure 2: (a) Second heart sound ensemble average of a patient without pulmonary arterial hypertension (NPAH). (b) Second heart sound ensemble average of a patient with PAH.

spline interpolation) were obtained for each S2, and the ensemble averaging applied to the positive and negative envelopes (*EnvMax* and *EnvMin*). This approach was inspired in the physiological knowledge that the S2 split in the PAH patients is increased, leading to well separated and audible A2 and P2 components, with an increased amplitude of P2 in relation to A2. This is observed in the PCG, and presented as a motivation for the extraction of the S2 envelopes. For each patient the ensemble averages of *EnvMax* and *EnvMin* were obtained ($EnvMax_\mu$ and $EnvMin_\mu$), as well as the difference between the two ($EnvDiff_\mu$); a peak-picking algorithm was applied to all envelopes to detect its two largest local maximums (A2 and P2 estimations, assuming A2 occurring first), and these amplitudes used in the estimation of the P2/A2 relation.

4 Results and Discussion

The best results in separating NPAH from PAH patients was obtained using $EnvMax_\mu$ (Wilcoxon Mann-Whitney test): $EnvDiff_\mu$ Total: 0.43±0.36, NPAH: 0.43±0.40, PAH: 0.42±0.18 (p=0.22); $EnvMin_\mu$ Total: 0.63±0.46, NPAH: 0.59±0.49, PAH: 0.74±0.34 (p=0.10); $EnvMax_\mu$ Total: 0.30±0.14, NPAH: 0.26±0.12, PAH: 0.42±0.10 (p<0.01).

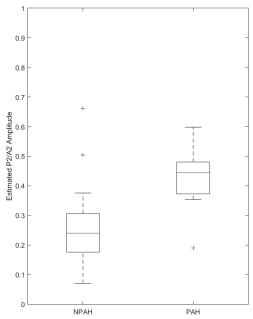


Figure 3: Boxplot of the P2/A2 ratio extracted from the maximum envelope ($MaxEnv_\mu$).

Figures 2(a) and 2(b) show the aligned S2s for one patient in each study group, the ensemble average of the S2s, envelopes, and the $MaxEnv_\mu$ detected peaks (estimated A2 and P2). It may be observed that the ensemble average of envelopes provides an estimate of the S2 energy spread within the respiration cycle. Figure 3 shows the amplitude relation between A2 and P2 (P2/A2) obtained from the $MaxEnv_\mu$; the P2/A2 ratio allows to separate the patients with and without PAH, and it may be observed that in the NPAH group the P2 amplitude rarely surpasses 50% of the A2 amplitude.

This algorithm has only been applied to a small sample. Larger clinical studies need to be conducted to evaluate the reproducibility of our results. But if proved reliable, this simple algorithm may be easily implemented in a mobile application, and has the potential to allow for a cost-effective tool to detect PAH by S2 shape analysis.

Acknowledgements

This work is funded by Instituto de Telecomunicações in the scope of project Rheumus, Ref. QREN 38505 and Ref. UID/EEA/50008/2013; by the Fundação para a Ciência e Tecnologia under the references Heart Safe PTDC/EEI-PRO/2857/2012; by Project I-CITY - ICT for Future Health/Faculdade de Engenharia da Universidade do Porto, funded by the Fundo Europeu de Desenvolvimento Regional (FEDER) through the Programa Operacional do Norte (ON2) and by national funds through FCT/MEC (PIDDAC), NORTE-07-0124-FEDER-000068.

References

[1] Ana Castro, Tiago T.V. Vinhoza, Sandra S. Mattos, and Miguel T. Coimbra. Heart sound segmentation of pediatric auscultations using wavelet analysis. In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pages 3909–3912, July 2013. doi: 10.1109/EMBC.2013.6610399.

[2] Mohamed Elgendi, Prashant Bobhate, Shreepal Jain, Long Guo, Jennifer Rutledge, Yashu Coe, Roger Zemp, Dale Schuurmans, and Ian Adatia. Spectral analysis of the heart sounds in children with and without pulmonary artery hypertension. *International journal of cardiology*, 173(1):92–99, 2014.

[3] Pedro Gomes, Samuel Frade, Ana Castro, Ricardo Cruz-Correia, and Miguel Coimbra. A proposal to incorporate digital auscultation and its processing into an existing electronic health record. In *HEALTHINF âĂŞ BIOSTEC 2015, 8th International Conference on Biomedical Engineering Systems and Technologies*, pages 143–150, Jan 2015. doi: 10.5220/0005222901430150.

[4] Mangione S and Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: A comparison of diagnostic proficiency. *JAMA*, 278(9):717–722, 1997. doi: 10.1001/jama.1997.03550090041030. URL `+http://dx.doi.org/10.1001/jama.1997.03550090041030`.

[5] Robert Smith and Dan Ventura. A general model for continuous non-invasive pulmonary artery pressure estimation. *Computers in Biology and Medicine*, 43(7):904 – 913, 2013. ISSN 0010-4825. doi: http://dx.doi.org/10.1016/j.compbiomed.2013.04.010.

[6] Morton E. Tavel. Cardiac auscultation: A glorious past - and it does have a future! *Circulation*, 113(9):1255–1259, 2006. doi: 10.1161/CIRCULATIONAHA.105.591149.

[7] Jeffrey M. Testani, Martin G. St. John Sutton, Susan E. Wiegers, Amit V. Khera, Richard P. Shannon, and James N. Kirkpatrick. Accuracy of noninvasively determined pulmonary artery systolic pressure. *The American Journal of Cardiology*, 105(8):1192 – 1197, 2010. ISSN 0002-9149. doi: http://dx.doi.org/10.1016/j.amjcard.2009.11.048.

[8] Jingping Xu, L.-G. Durand, and P. Pibarot. Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model. *Biomedical Engineering, IEEE Transactions on*, 48(3):277 –283, march 2001. ISSN 0018-9294. doi: 10.1109/10.914790.

[9] Ya-Li Zheng, Xiao-Rong Ding, C.C.Y. Poon, B.P.L. Lo, Heye Zhang, Xiao-Lin Zhou, Guang-Zhong Yang, Ni Zhao, and Yuan-Ting Zhang. Unobtrusive sensing and wearable devices for health informatics. *Biomedical Engineering, IEEE Transactions on*, 61(5):1538–1554, May 2014. ISSN 0018-9294. doi: 10.1109/TBME.2014.2309951.

Evaluation of cell adhesion status through quantitative analysis of DAPI staining fluorescence images

Tânia Mestre[1]
tania.raquel.mestre@tecnico.ulisboa.pt
Raquel Seruca[2]
rseruca@ipatimup.pt
João M. R. Sanches[1]
http://users.ist.isr.utl.pt/~jmrs

[1] Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisboa, Portugal.

[2] Instituto de Patologia e Imunologia Molecular da Universidade do Porto
Porto, Portugal.

Abstract

Cell adhesion is a crucial mechanism in the formation and maintenance of three-dimensional patterns of cells in epithelial tissues. Loss of cell adhesion increases the risk for cell invasion of the surrounding tissues and metastasis. Furthermore, alterations in the epithelial protein E-cadherin, which is the major contributor to these phenomena, are typically associated with invasive behavior of several types of carcinomas, the stomach as example.

In this work, quantitative imaging tools are proposed to infer the adhesion status of cells with mutant E-cadherin expression through the use of microscopic images of cells stained with 4',6-diamidino-2-phenylindole (DAPI). The computational-based quantification of cell adhesion loss was pursued through the analysis of the intercellular network features attained from these images.

1 Introduction

In recent years, advances in the development of quantitative methods to evaluate cellular adhesion have been established namely, AFM (Atomic Force Microscopy) - based Single-cell Force Spectroscopy [1][2] and FRET (Förster Resonance Energy Transfer) - based molecular Tension Sensitive Module [1][3], among others. These tools provide the possibility to identify mechanisms underlying the formation and maintenance of intercellular contact, cell migration as well as tissue remodeling [1]. However, these methods present several limitations, such as their inherent high costs and necessity of direct manipulation of cells.

This work aims to quantify the intercellular distance and cellular organization to infer the status of cell adhesion of distinct cell lines, harboring wild type or E-cadherin mutations (A634V, R749W and E781D), through DAPI stained fluorescence microscopy images. In order to perform this analysis, several features (areas, variation of distances and angles between neighboring cells) were evaluated regarding the intercellular network of distinct cellular cultures.

2 Proposed Method

Fluorescence microscopy (FM) images obtained from cells in culture expressing diverse forms of E-cadherin were the basis of this work. Only the information retrieved from the DAPI (blue) plane was used and a pre-processing pipeline of the FM images was pursued. This pipeline consisted on the application of a 1) denoising strategy, developed by Rodrigues et al. (2011) [4], in order to remove the Poisson intensity dependent noise, 2) a contrast [5] and intensity [6] adjustment performed to each image, in order to obtain more distinguishable objects, 3) the nuclei segmentation process, based on the application of the Otsu's method [7], morphological operators [8][10] and seeded watershed algorithm [9][10][11], 4) the regions segmentation process, based on the Otsu's method [7] and watershed algorithm [11] and finally 5) the definition of neighbor nuclei, based on the output of steps 4 and 5, where two nuclei were considered neighbors if the total number of label transitions was inferior or equal to 4. These steps therefore allowed a semi-automatic attainment of an intercellular network representative of each FM image, strictly composed by triangles.

The pre-processing pipeline applied therefore allows us to obtain an intercellular network, defined through the nodes of the intercellular triangular network, which represent the geometrical centers of the cells, and the edges, which represent the neighboring relations between cells, where triplets of neighboring cells are associated in each network triangle and each cell may belong to more than one triangle. An example of these intercellular networks is shown in Fig. 1.

In order to characterize the intercellular network of each cell line in terms of size of the triangles and also heterogeneity, several features were extracted, namely, the areas, as well as the variation of distances ($\gamma_{distance}$) and angles (γ_{angle}) of each triangle constituting the network, identified schematically in Fig. 2.

$$\text{Area} = \frac{1}{2}\left|\overrightarrow{AB} \times \overrightarrow{BC}\right| \tag{1}$$

$$\gamma_{distance} = \frac{\frac{1}{N}\sum_{i=1}^{N}(d_i - \mu_d)^2}{\mu_d{}^2} \tag{2}$$

$$\gamma_{angle} = \frac{1}{N}\sum_{i=1}^{N}(a_i - \mu_a)^2 \tag{3}$$

in which N = 3, since it represents the total number of distances and angles of each triangle, d_i and a_i are the distances and angles of each triangle and μ_d and μ_a the respective means.

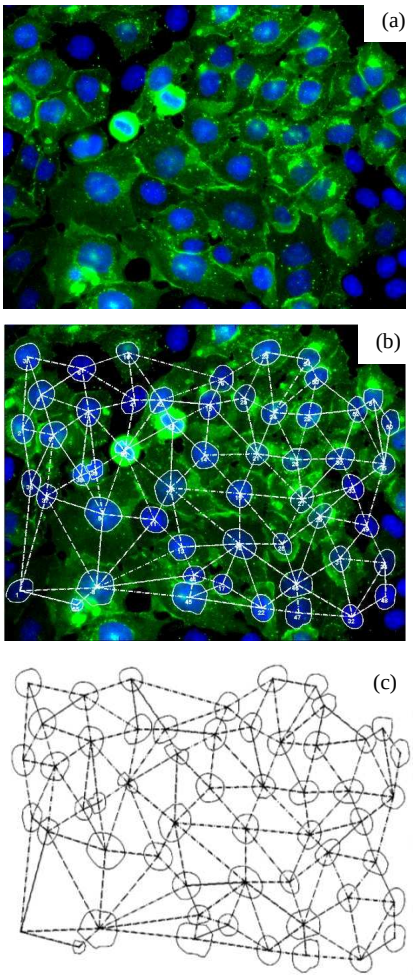


Figure 1: (a) Fluorescence microscopy image originated from cellular cultures of Chinese Hamster Ovary (CHO) cells transfected with vector encoding the wild type human E-cadherin. (b) Image (a) overlapped with the corresponding intercellular network. (c) Intercellular network of image (a).

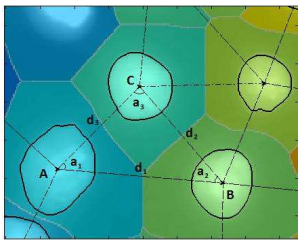


Figure 2: Centroids (A, B, C), distances (d1, d2, d3) and angles (a1, a2, a3) of each triangle constituting the triangular network, not necessarily in the order specified in the figure.

3 Experimental Results

All biological replicas of a given cell line (wild type or mutant), instead of each image, were assembled in all features analyzed. The results were plotted in normalized histograms in order to obtain a probability density function. A comparison between mutated and non-mutated cells was performed by overlapping the histograms. Each green histogram in Fig. 3 refers to a distinct E-cadherin mutation, whose affected domain is identified in the upper scheme of the same figure. A statistical analysis of intercellular network results, based on the Mann-Whitney U test [12], was subsequently obtained in order to verify statistically significant differences between cell lines with normal expression of E-cadherin and cell lines with aberrant expression of this protein.

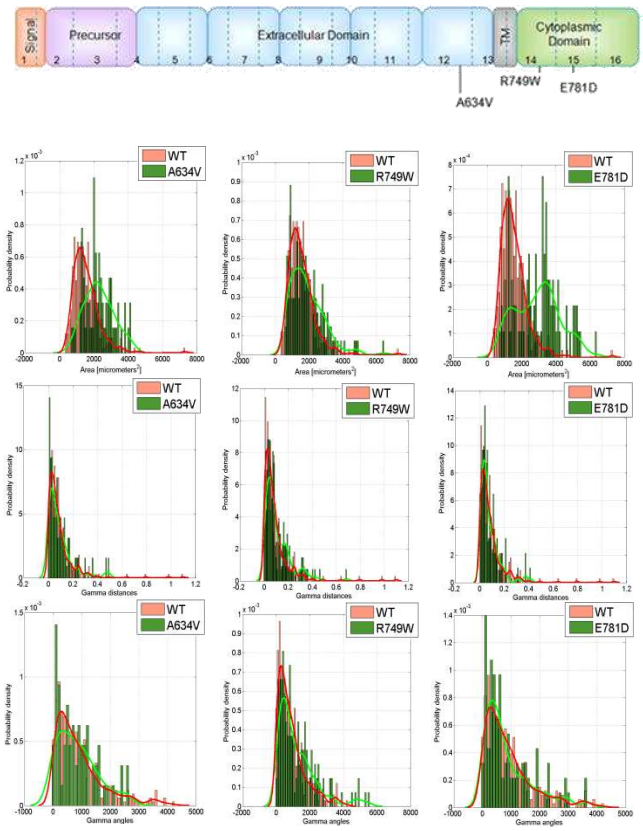


Figure 3: Probability density function of the triangular network areas, $\gamma_{distance}$ and γ_{angle} metrics obtained from cellular cultures of CHO cells transfected with vectors encoding the wild type (in red) and mutant variants of human E-cadherin (in green). The curves in red and green represent the probability density estimate, based on a normal kernel function.

The obtained results showed that all mutated cell lines presented higher mean and standard deviation values of the network areas in comparison to the wild type cell line. In all cases with mutations, the results obtained were statistical significant. Therefore, the areas of the triangles extracted from the triangular network reflect loss of cell-cell adhesion in mutated cell lines.

Interesting, we verified that the analysis of both $\gamma_{distance}$ and γ_{angle} metrics were accurate measures to evaluate the intercellular network heterogeneity in some cell lines. In R749W mutant cells, $\gamma_{distance}$ and γ_{angle} results are associated to more irregular meshes representing lower intercellular organization.

Therefore, these findings suggest that mutant E-cadherin cells besides being more distant from each other due to loss of cell adhesion, the intercellular organization is also perturbed by the presence of abnormal expression of E-cadherin when compared to cells line with normal expression of this protein.

4 Conclusions

The proposed methodology disclosed statistically significant results regarding the distinction of mutated from non-mutated cell cultures through the analysis of the intercellular networks attained from DAPI staining FM images.

The major advantage of the proposed technique regarding several others is the possibility to screen through an easy, costless and fast process mutated cell lines, using quantitative morphological parameters that can infer the cell adhesion status. The biological intercellular networks evaluated revealed significant results that would enable one to easily relate the analyzed features with the weakening of cellular adhesion, already verified through distinct *in vitro* tests for the mutations in study.

References

[1] J. Kashef and C. M. Franz, "Quantitative methods for analyzing cell–cell adhesion in development," Developmental biology, 2014.

[2] Fichtner, Dagmar, et al. "Covalent and density-controlled surface immobilization of E-cadherin for adhesion force spectroscopy." *PloS one* 9.3 (2014).

[3] Borghi, Nicolas, et al. "E-cadherin is under constitutive actomyosin-generated tension that is increased at cell–cell contacts upon externally applied stretch." *Proceedings of the National Academy of Sciences* 109.31 (2012): 12568-12573.

[4] C. Rodrigues and J. M. R. Sanches, "Convex total variation denoising of poisson uorescence confocal images with anisotropic filtering," Image Processing, IEEE Transactions on, vol. 20, no. 1, pp. 146-160, 2011.

[5] Mathworks, "adapthisteq", Available from: "http://www.mathworks.com/help/images/ref/adapthisteq.html", accessed on 08.04.2015.

[6] Mathworks, "imadjust", Available from: "http://www.mathworks.com/help/images/ref/imad just.html", accessed on 08.04.2015.

[7] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms."*Automatica* 11.285-296 (1975): 23-27.

[8] J. C. Russ, "The image processing handbook", Second edition. IEEE press, 1995.

[9] Wählby, Carolina, et al. "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections." *Journal of Microscopy* 215.1 (2004): 67-76.

[10] Lin, Gang, et al. "A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks." *Cytometry Part A* 56.1 (2003): 23-36.

[11] Pinidiyaarachchi, Amalka, and Carolina Wählby. "Seeded watersheds for combined segmentation and tracking of cells." *Image Analysis and Processing–ICIAP 2005*. Springer Berlin Heidelberg, 2005. 336-343.

[12] L. Statistics, "Mann-whitney u test using spss," Available from: "https://statistics.laerd.com/ spss-tutorials/mann-whitney-u-test-using-spss-statistics.php", accessed on 17.04.2015.

Metacarpal & Phalange Bones Segmentation from Ultrasound Images

Malik Saad Sultan [1,2]
engr.saadsultan@gmail.com
Nelson Martins [1,3]
nelsonmartins89@gmail.com
Manuel Joao Ferreira [3]
mferreira@enermeter.pt
Miguel Tavares Coimbra [1,2]
mcoimbra@dcc.fc.up.pt

[1] Instituto de Telecomunicações
[2] Faculdade de Ciências, Universidade do Porto, Portugal
[3] Enermeter, Sistemas de Medição, Lda

Abstract

The accurate segmentation of the bone surface and the identification of the MCP capsule region remains a challenge in ultrasound image processing. In this article we aim to make a contribution to this problem by incorporating prior knowledge of the bone and joint regions anatomy into our segmentation algorithm. The log Gabor filter is used for speckle noise reduction and to extract ridge-like structures from the images, while the phase is left unchanged. After thresholding, scores are generated, based on the intensities and areas of the resulting regions, enabling the selection of the structure that best matches the bone. Finally, segmented joint bones are processed to calculate the initial seeds of the joint capsule region. Experimental results demonstrate the accuracy of the proposed segmentation algorithm. The mean pixel error between the automatic segmentation and the reference images were 4.4 pixel. The bone regions not segmented were, on average, 5.4%.

1 Introduction

Rheumatic arthritis is the main cause of impairment and pain in developed countries, which makes them a critical social, health and economic problem. Ultrasound is one of the recommended techniques for the assessment of inflammatory activities in soft tissues and in the bone erosion process, especially in the small joints of RA patients [1, 2]. The difficulties in the interpretation and acquisition of this kind of images, reduces its acceptance, which motivates the research and development of computer assisted decision solutions to help physicians.

Ultrasound images are usually degraded by strong speckle noise, which appears as grain of different sizes and intensity. Wavelet shrinkage denoising is widely used for speckle noise reduction. As an example, wavelets based on complex valued log Gabor functions can be used to estimate noise in each scale. The noise threshold is estimated in each scale and then the magnitude of the filter vector response is calculated, while the phase is left unchanged [3].

Several methods were proposed to automatically extract bone surface [4, 5] and to quantify synovial inflammation from ultrasound images [6, 7]. The 2D log Gabor filter is used to estimate phase symmetry features from ultrasound images for the segmentation of bone surfaces [5].

In this work we propose a new bone segmentation method. We use prior knowledge of the expected bone appearance in an ultrasound image. A new feature is introduced, that is the area covered by the metacarpal and phalange bone. The algorithm is designed which combine this feature with previously known features (intensity, shadow). An algorithm is proposed to address the intensity drop-off problem at the joint. Initial seeds were estimated to roughly segment the MCP joint region. The observed anatomical structures in an ultrasound images are illustrated in Fig. 1.

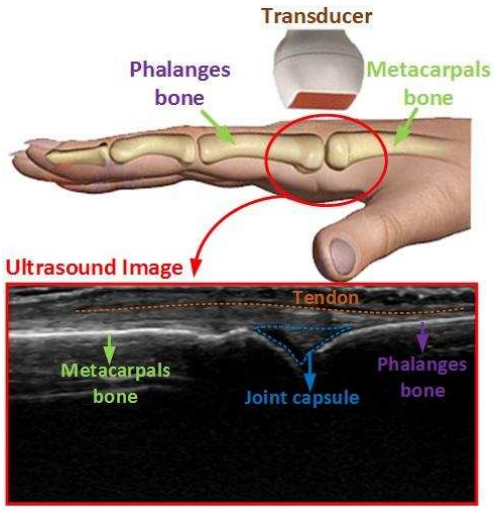


Figure 1: Observed anatomical structures

The structure of the paper is organized as follows. Section II provides the methodology adopted in this paper. In section III we report the results that demonstrate the accuracy of the proposed algorithm and finally section IV concludes the paper with a discussion on the problem and our contribution to it.

2 Methodology

The work flow diagram of the full algorithm is shown in Fig. 2

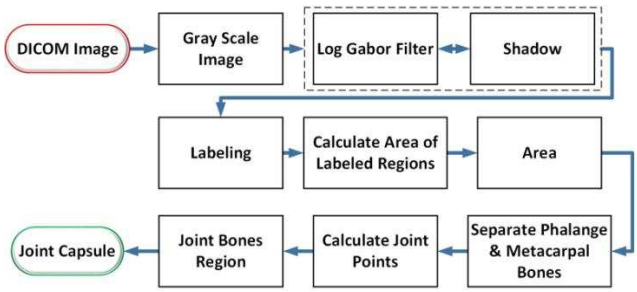


Figure 2: Overview of the Algorithm

2.1 Pre-processing

The log Gabor wavelets are efficient to suppress noise components from the noise [2]. Complex valued log Gabor wavelets are used that preserve the phase information. Denosing can be considered as a three-step process.

In the first step, the phase and amplitude information is calculated using the symmetric pair of continuous wavelet and log Gabor filters in quadrature at each scale.

The second step determines the noise threshold. The magnitude of the filter response vector is estimated based on the calculated noise threshold at each scale. In this approach the shrinkage of the complex (real and imaginary part) valued wavelets response vector is considered.

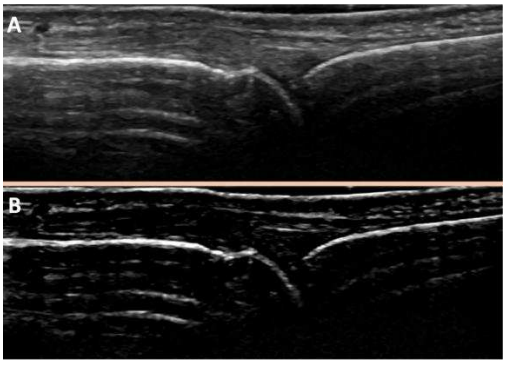


Figure 3: A) Original image B) Denoised Image

The smallest scale n is used because it holds the strongest noise response and the largest bandwidth. Finally, in each scale of the orientations, the value that exceeds the defined threshold is discarded as noise (See Fig. 3).

2.2Bone segmentation

In the proposed algorithm, the bones are segmented based on their anatomy, which creates a specific response in ultrasound images [6]. As the bones are continuous structures, they produce long ridges with some specific thickness (regions with large areas).

The top part of the image is cropped. This includes soft tissue and is very close to the transducer, which means no bones will be visible here. The intensity and shadow of the bone are used as a constraint to minimize the false positive and false negative detections. The results are multiplied by the filtered image to estimate the probability of the presence of a bone in an image pixel. Morphological operations are used to remove small discontinuities between close bone regions. All the detected regions are labeled and then the area, centroid and major axis length are calculated for each one. Finally, label areas are sorted by descending order and the two biggest regions are extracted. At this point we found that, the probability of these regions to be phalange and metacarpal is typically very high.

The medial axis transform simplifies the shape of both segmented bones. The joint points are separated and used to detect regions that are closer to the bone. The algorithm separates the region of interest by using the joint points of metacarpal and phalange bone. The close regions are segmented in an iterative way. The distances between the metacarpal joint point and centroids of each region are calculated. Based on this distance, the closest region is selected and considered as the part of the bone. The selected bone centroid is then considered as an initial point and this procedure continues until other region is close enough. In this way the whole bone surface in an ultrasound image can be segmented.

3 Result

3.1Material

The proposed algorithm was tested on fifteen images, acquired with a GE Healthcare LOGIQ-S8. Images are from the MCP joint of the index finger with a size of (488\times761) in DICOM format and with a spatial resolution of 0,0531mm per pixel. The proposed algorithm was implemented with MATLAB R2013a, in a window-7 operation system. In order to validate our proposed algorithm the images were manually segmented by an expert.

3.2Bone segmentation results

Figure 4 shows the obtained results with the proposed algorithm. The expert annotation is shown with red color and the automatic segmentation results are shown in green (Fig. 4).

In the Fig. 4b it is clearly visible that the expert annotation is mostly overlapped by the automatic segmentation, which strengthens the belief that this is a viable approach to address this problem. The mean pixel error between the expert labeled data and the automatic segmentation is calculated in the Y-axis. In order to quantify the regions where the algorithm failed to segment the bone, we measure the bone region percentage not segmented. The mean, maximum, minimum and standard deviations were calculated (see Table. 1). The best results in terms of false positive and negative (soft tissues as bone, vice versa) were obtained in the clear images with detection rates of about 97%. Lower image quality compromises results since it may identify some irrelevant regions as a bone because of similar properties (intensity, area). Future work needs to improve this robustness to noise.

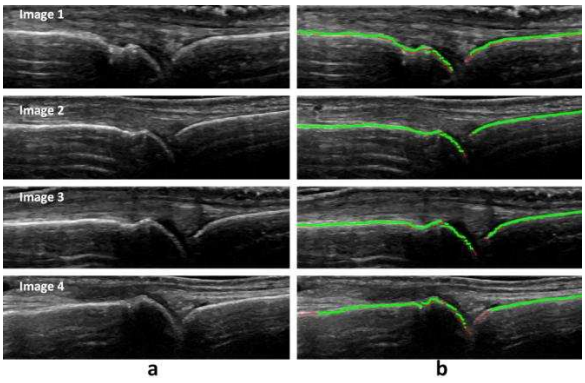


Figure 4: Segmentation results of US image. (a) Original images (b) Red and green colors represent manual and automatic segmentation, respectively

Table 1: Segmentation error

Mean pixel error per exam (Pixel)			
Mean: 4.4	Max: 5.6	Min: 3.7	Std Dev: 0.6
Bone region percentage not segmented (%)			
Mean: 5.40%	Max: 23%	Min: 1%	Std Dev: 5.50%

4 Discussion

Results show that it is possible to segment bone and joint regions in metacarpophalangeal ultrasound images with reasonable accuracy, leading to the automatic estimation of relevant physiological parameters for computer-assisted decision systems. These initial results are obtained with a small set of images, and studies are underway to validate these in a larger dataset.

The proposed algorithm shows that it is possible to achieve results similar to the ones obtained by an expert. Regarding the mean pixel error, results are very close an expert. The offsets measured are relatively small, around 3.7-5.6 pixel, and missing rates, were satisfactory; with minimum of 1% and maximum of 23% pixels missed.

Acknowledgement

This work is funded by Instituto de Telecomunicações in the scope of Project UID/EEA/50008/2013 and by Fundação para a Ciência e Tecnologia (FCT) grant no: PD/BD/105761/2014

References

[1] A. Ficjan et. al., Ultrasound composite scores for the assessment of inflammatory and structural pathologies in Psoriatic Arthritis, Arthritis Research & Therapy, vol. 16, Oct 31, 2014

[2] P. Zufferey et. al., Recommendations for the use of ultrasound in rheumatoid arthritis: literature review and SONAR score experience, Swiss Med Wkly, Dec 20, 2013

[3] D. J. Field et. al., Relations between the statistics of natural images and the response properties of cortical cells, Journal of the optical society of america, vol. 4, pp. 2379-2394, Dec 1987

[4] P. Foroughi et. al., Ultrasound Bone Segmentation Using Dynamic Programming, Ultrasonics Symposium, pp. 2523-2526, Oct 2007

[5] I. Hacihaliloglu et. al., Automatic Data-Driven Parameterization for Phase-Based Bone Localization in US Using Log-Gabor Filters, Advances in Visual Computing Lecture Notes in Computer Science, Vol. 5875, pp. 944-954, 2009

[6] K. A. Patwardhan et. al., Automated bone and joint-region segmentation in volumetric ultrasound, International Symposium on Biomedical Imaging , pp. 1327- 1330, May 2012

[7] D. M. Mills et. al., Volumetric Ultrasound and Computer-Assisted Analysis At The Point-of-Care: A Musculoskeletal Exemplar, 34th IEEE Annual International Conference of the EMBS, pp. 2318-2322, 2012

Using Probabilistic Logic Programming to Find Patterns

Theofrastos Mantadelis[1]
theo.mantadelis@dcc.fc.up.pt
Ricardo Rocha[1]
ricroc@dcc.fc.up.pt
Jorge Oliveira[2]
oliveira_jorge@dcc.fc.up.pt
Miguel Tavares Coimbra[2]
mcoimbra@dcc.fc.up.pt

[1] CRACS & INESC TEC
Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre 1021/1055,
4169-007 Porto, Portugal

[2] Instituto de Telecomunicações
Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre 1021/1055,
4169-007 Porto, Portugal

Abstract

This short paper, briefly presents the probabilistic logic programming language ProbLog and the system MetaProbLog. We present an example Hidden Markov Model to illustrate the three main tasks of the system. Furthermore, we mention some of the existing ProbLog applications which are used to find connections/patterns in relational databases. Finally, we present an application that uses MetaProbLog for phonocardiogram classification which is used in order to diagnose heart diseases.

1 Introduction

Probabilistic Logic Programming (PLP) combines technologies from logic programming, knowledge representation and reasoning and machine learning. Probabilistic models in our days are principled and a widely used approach to deal with uncertainty. First order logic can elegantly represent complex situations involving a variety of objects as well as relations among the objects.

MetaProbLog[1] [6] is a framework of the ProbLog [3, 5] probabilistic logic programming language. ProbLog extent Prolog programs by annotating facts with probabilities. In that way it defines a probability distribution over all Prolog programs. ProbLog follows the distribution semantics presented by Sato [9]. MetaProbLog extends the semantics of ProbLog by defining a "ProbLog engine" which permits the definitions of probabilistic meta calls [6]. MetaProbLog inference, currently allows the computation of marginal probabilities with or without evidence. Furthermore, it allows the computation of marginal probabilities for the answers of non-ground queries.

MetaProbLog has three primary inference methods: exact inference, program sampling and most probable explanation. The exact inference method, uses state of the art knowledge compilation methods [2]; program sampling, is a rejection sampling approach; and finally, the most probable explanation inference uses a dynamic algorithm to find the most probable explanation of a query.

2 Semantics

A ProbLog program T consists of a set of facts annotated with probabilities $p_i :: pf_i$ – called *probabilistic facts* – together with a set of standard definite clauses $h : -b_1, \ldots, b_n$. that can have positive and negative probabilistic literals in their body. A probabilistic fact pf_i is true with probability p_i. These facts correspond to random variables, which are assumed to be mutually independent. Together, they thus define a distribution over subsets of $L_T = \{pf_1, \ldots, pf_n\}$. The definite clauses add arbitrary *background knowledge* (BK) to those sets of *logical* facts. To keep a natural interpretation of a ProbLog program we assume that probabilistic facts cannot unify with other probabilistic facts or with the background knowledge rule heads. Formally, a ProbLog program is of the form $T = \{pf_1, \ldots, pf_n\} \cup BK$.

Given the one-to-one mapping between ground definite clause programs and Herbrand interpretations, a ProbLog program defines a distribution over its Herbrand interpretations.

The distribution semantics are defined by generalising the least Herbrand models of the clauses by including subsets of the probabilistic facts. If fact pf_i is annotated with p_i, pf_i is included in a generalised least Herband model with probability p_i and left out with probability $1 - p_i$. The different facts are assumed to be probabilistically independent, however, negative probabilistic facts in clause bodies allow the user to enforce a choice between two clauses.

As such, a ProbLog program specifies a probability distribution over all its possible non-probabilistic subprograms. The success probability of a query is defined as the probability that the query succeeds in such a random subprogram. ProbLog follows the distribution semantics [9] proposed by Sato.

[1] MetaProbLog's website: `www.dcc.fc.up.pt/metaproblog`

3 Example Program & Queries

The syntax of MetaProbLog uses logic programming, specifically Prolog, in order to be very expressive as a language and be able to describe complex models. Next we present a small MetaProbLog program that defines the Hidden Markov Model illustrated at Figure 1, the different colors indicate the possible transitions from each state which are modeled by **annotated disjunctions**[2] in the program.

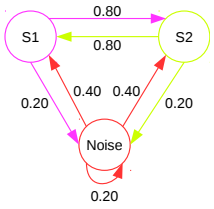


Figure 1: The graphical representation of a 3 state Hidden Markov Model

Below follows the MetaProbLog program that models the Hidden Markov Model of Figure 1.

```
0.80::trans(s1,s2,T1,T2);
0.20::trans(s1,noise,T1,T2) <- next(T1, T2).

0.80::trans(s2,s1,T1,T2));
0.20::trans(s2,noise,T1,T2) <- next(T1, T2).

0.40::trans(noise,s1,T1,T2);
0.40::trans(noise,s2,T1,T2);
0.20::trans(noise,noise,T1,T2) <- next(T1, T2).

0.20::start(noise,0);
0.40::start(s1,0);
0.40::start(s2,0) <- true.

signal(State, 0) :- start(State, 0).
signal(State2, T2) :-
  trans(State1,State2,T1,T2),
  signal(State1, T1).

next(T1, T2) :- integer(T1), !, T2 is T1 + 1.
next(T1, T2) :- integer(T2), T1 is T2 - 1.
```

For our example we use the term `trans/4` to describe a transition of the model from one state to another (first and second argument of the

[2] ProbLog's syntax: `P1::Choice1 ; ⋯ ; PN::ChoiceN <- Body`, is used to model exclusive choices with $\sum P1 \cdots PN = 1.0$. This construct is called annotated disjunction and two choices of an annotated disjunction can never be true at the same time.

term), at a time to the next time (third and fourth argument of the term). By composing an annotated disjunction with the appropriate `trans/4` terms we compactly describe all the transitions our model can take. Furthermore, the term `start/2` is used to define the possible starting states. Finally, the predicate `signal/2` defines the infinite chain and dependencies of the model.

MetaProbLog can ask queries about the probability distribution of such a model. For example: we can ask what is the probability that at a specific time we have a specific state (Query 1); we can ask what is the conditional probability of a specific state using some prior knowledge like the starting state of the model (Query 2); and finally, we can query what is the most likely set of states that reach to a specific result (Query 3).

Below we present the three example queries and their results.

```
% Query 1: Probability of a state
   ?- problog_exact(signal(s1,5),P).
P = [0.4]

% Query 2: Conditional probability of a state
%          with prior knowledge
   ?- problog_exact(signal(s1,5)/start(s1,0),P).
P = [0.23616]

% Query 3: Most probable explanation of a query
   ?- problog_mpe(signal(s1,5),Res).
Res = [0.131072/[start(s2,0)->true,
                 trans(s2,s1,0,1)->true,
                 trans(s1,s2,1,2)->true,
                 trans(s2,s1,2,3)->true,
                 trans(s2,s1,4,5)->true,
                 trans(s1,s2,3,4)->true]]
```

While the ProbLog language was introduced to answer statistical questions for relational models one can easy observe that the same question could be addressed on patterns. For example: Query 3 could be asking what is the most probable pattern of the model.

MetaProbLog is able to model all statistical graphical models such as Hidden Markov Models, Bayesian Networks, Probabilistic Graphs. Furthermore, any Prolog program could be extended with MetaProbLog to use probabilities and take decisions with them.

4 Applications

ProbLog systems have found applications in many fields with most common examples to include:

- Link discovery in Biomine Alzheimer database [11]. Biomine Alzheimer database is a real-world biological dataset of Alzheimer genes which corresponds to a directed probabilistic graph of 11530 edges and 5220 nodes. ProbLog was used to discover relations among genes and other biological properties [3].
- WebKB (`http://www.cs.cmu.edu/~webkb`) is a dataset from a collective classification domain in which university webpages are classified according to their textual content. ProbLog has been used to learn the probabilities that two webpages are related and to query the WebKB [4].
- The probabilistic Dictionary [12] is used to discover the probability that two words have the same meaning. It includes around 250 different words from the English language and meanings for about 30 of them. Some words are related together according to their semantic relatedness. This relation is marked with the probability that the two words have the same meaning.
- ProbLog has also been used for robotic affordance model learning by [8]. In this application ProbLog was successfully used to learn a robotic task with multiple objects and complex spacial relations.
- Finally, ProbLog has been used to model Mobile Ad hoc Networks and analyse Fadip [7], a Publish/Subscribe protocol for Mobile Ad hoc Networks. ProbLog efficiently calculates the probability that a message would be transmitted from one device to another in the network, analysing statistics for the traffic and reachability of the protocol.

Lately, the classification of phonocardiogram (PCG) signals has got significant attention in the academic community [1]. Classifying PCGs is both a challenging and an important task. Heart sounds are non-trivial signals, since they might contain non-stationary noise, have artifacts and murmur sounds. Heart sound auscultation techniques is one of the most reliable and successful tools in early diagnosis used for potentially deadly heart diseases, such as natural and prosthetic heart valve dysfunction or even in heart failure. Therefore a computer-aided auscultation may allow detection of diseases that are hardly recognized through the traditional methods, for instance ischemic heart disease.

Recently, HMMs have being used for modeling and characterizing real-world signals such as heart sound signals [10]. For future work, we aim to model PCG signals as a HMM and use MetaProbLog to find the most likely sequence of events (S1, S2, S3, S4, noise, murmur, etc.) and finally, use our model in order to characterize real life segmented signals.

Acknowledgements

This work is funded by the Instituto de Telecomunicações in the scope of Project UID/EEA/50008/2013 and the Fundação para a Ciência e a Tecnologia (FCT) (Portuguese Foundation for Science and Technology) within the project UID/EEA/50014/2013.

References

[1] P. Bentley, G. Nordehn, M. Coimbra, and S. Mannor. The PASCAL classifying heart sounds challenge 2011 results. `http://www.peterjbentley.com/heartchallenge`.

[2] A. Darwiche and P. Marquis. A knowledge compilation map. *Journal of Artificial Intelligence Research*, 17:229–264, 2002.

[3] L. De Raedt, A. Kimmig, and H. Toivonen. ProbLog: a probabilistic Prolog and its application in link discovery. In *International Joint Conferences on Artificial Intelligence*, pages 2468–2473, 2007.

[4] D. Fierens, G. Van Den Broek, J. Renkens, D. Shterionov, B. Gutmann, I. Thon, G. Janssens, and L. de Raedt. Inference and learning in probabilistic logic programs using weighted boolean formulas. *Theory and Practice of Logic Programming*, 15(03):358–401, 2015.

[5] A. Kimmig, B. Demoen, L. De Raedt, V. Santos Costa, and R. Rocha. On the implementation of the probabilistic logic programming language ProbLog. *Theory and Practice of Logic Programming*, 11:235–262, 2011.

[6] T. Mantadelis and G. Janssens. Nesting probabilistic inference. *Computing Research Repository*, abs/1112.3785, 2011.

[7] T. Mantadelis, K. Paridel, G. Janssens, Y. Vanrompay, and Y. Berbers. Analysing a Publish/Subscribe System for Mobile Ad Hoc Networks with ProbLog. In *Practical Aspects of Declarative Languages*, pages 34–37, 2011.

[8] B. Moldovan, P. Moreno, M. van Otterlo, J. Santos-Victor, and L. De Raedt. Learning relational affordance models for robots in multi-object manipulation tasks. In *International Conference on Robotics and Automation*, pages 4373–4378, May 2012.

[9] T. Sato. A statistical learning method for logic programs with distribution semantics. In *International Conference on Logic Programming*, pages 715–729, 1995.

[10] P. Sedighian, A.W. Subudhi, F. Scalzo, and S. Asgari. Pediatric heart sound segmentation using Hidden Markov Model. In *Engineering in Medicine and Biology Society*, pages 5490–5493, Aug 2014.

[11] P. Sevon, L. Eronen, P. Hintsanen, K. Kulovesi, and H. Toivonen. Link discovery in graphs derived from biological databases. In *International Conference on Data Integration in the Life Sciences*, pages 35–49, 2006.

[12] D. Shterionov and G. Janssens. Data acquisition and modeling for learning and reasoning in probabilistic logic environment. In *Portuguese Conference on Artificial Intelligence*, pages 298–312, 2011.

Beef Quality Evaluation System

Cátia Teixeira
a25433@alunos.ipb.pt
Vasco Cadavez
vcadavez@ipb.pt
Fernando C. Monteiro
monteiro@ipb.pt

Polytechnic Institute of Bragança, ESTiG,
Bragança, Portugal

Abstract

Applying computer vision in meat quality evaluation has been an active area of research in recent years. Accurate segmentation of beef-marbling images plays an important role in making the correct decision on beef-marbling score in an automatic beef quality grading system. The purpose of this study is to develop a new segmentation method to correctly separate the fat flecks from the muscle in the rib-eye region in a beef image. This paper presents an automatic system for beef marbling measuring which is composed of discriminant threshold selection method and run length processing. From the experimental results, it has been confirmed that the proposed system enables high quality grading of beef marbling, and robust region segmentation of the actual beef rib-eye image into lean and fat regions.

1 Introduction

The existing methods for meat quality classification rely heavily on visual evaluation of certain characteristics of the carcasses which according to the United States Department of Agriculture beef quality grading system, are the abundance of marbling, the colour of muscle and the skeletal maturity [7]. The features of *longissimus dorsi* (LD) colour, marbling and surface texture can be used to generate predictive models of eating quality parameters using classical statistical methods [3].

Marbling in beef is an important factor that influences taste, juiciness, tenderness and flavour [4] and it is often considered by consumers as the primary factor when buying meat in view of its contribution to sensory characteristics of meat [2]. Beef marbling is evaluated in terms of the content of fat and the distribution of fat particles. For current standards of grading beef quality, the beef marbling score is assigned from visual appraisal of the LD muscle. Generally a grader determines the grade of the marbling by visual inspection, which involves the collation of the actual carcass cross-section with published standard photographs illustrating marbling abundance of each grade. Although graders are professionally trained, subjective evaluation often lead to inconsistencies and variations.

The beef-marbling pattern is irregular, complex and varied. There is not clear bound between the lean region and the fat one. In spite of many researches accurate segmentation of beef marbling remains a challenging problem in automatically grading beef marbling by machine vision and image processing [1, 3, 6]. These evaluation methods do not replace the sensory evaluation, but allow a standardized methodology [2].

Chen and Qin [1] described a method for identifying the inter-muscular fat streaks by thresholding based on a between-class variance method. It performs well both in relatively light and also in relatively dark images, and this has shown to be more robust than eight standard algorithms, including region growing, the Otsu method and maximum entropy thresholding.

Jackman et al. [3] used Fuzzy C-means to split colour image into background, lean and fat. Then, they manually chose the threshold level to remove marbling pixels and eliminate any background pixels. All holes in LD image with 3 pixels or larger are considered as marbling.

Relatively good segmentation results have been obtained by the clustering method, which can provide the segmentation of fat flecks from lean muscle regardless the image histogram as proposed by Subbiah et al. in [6]. The original image was segmented into red lean muscle, white marbling and blue background. A number of images contained partially buried marbling flecks, which led to an overestimate of the LD muscle as the pink pixels were not classed as marbling. A thresholding operation was performed to remove those pixels from the LD image. The reflection of any connective tissue surrounding the LD of the muscle folds inwards were removed manually.

However, clustering analysis is an iteration process that has the disadvantages of being time-consuming and of low efficiency compared to the method based on thresholding. In order to reduce computation time, some efforts have been made to reduce the dimensionality of the original image. For instance, Subbiah et al. [6] reduced the image size by down sampling and Pang et al. [5] reduced the dimensionality of the original image by resampling.

In this paper we propose a computer vision framework that can be used in beef quality evaluation. The key idea is to measure the percentage of marbling in the muscle to obtain a beef quality evaluation system.

The paper is organized as follows: In “Materials and Methods,” the proposed computer vision framework and the experiments are explained. In “Results and Discussion,” the results obtained in several experiments are shown. Finally in “Conclusions” some concluding remarks are given.

2 Materials and Methods

The proposed framework used to design automatically a computer vision system for beef quality evaluation consists of five steps:

1. Image acquisition: A digital image of the beef under test is taken.
2. Pre-processing: The digital image is improved in order to enhance the details.
3. Segmentation: The region of interest (LD muscle) is found and isolated from the background of the scene.
4. Feature extraction/selection: Significant features of the beef are extracted in order to isolate the muscle and the marbling.
5. Classification: The extracted features are measured automatically in order to evaluate its quality.

2.1 Data Acquisition

Samples were obtained from carcasses in *Terra Fria Carnes* slaughterhouse. The materials included 83 cross sections in the standard location for measuring marbling.

All sample images were captured under the slaughter house artificial illumination. A removable pan with a white surface was used to position the steak in the camera field-of-view. The flash unit in the digital camera was not used because the light reflectance on meat surface lead to brightness errors.

2.2 Proposed Method

The process for checking beef marbling by image analysis can be largely divided into two sub-processes, based on the process of visual inspection performed by the graders. The initial process is to segment a cross-sectional image of beef carcass rib-eye into lean and fat regions. The second is to quantify features of beef marbling which are then used to estimate the beef quality.

The image pre-processing refers to the operation separating the rib-eye from the background. This is a necessary procedure for the segmentation of marbling, which is composed of several operations such as thresholding and morphological operation. As the images have a very complex background, it was decided to make a cut on the original images in order to reduce most objects of the background obtaining simpler and easier images to segment.

Since there are different light conditions among image acquisition tasks, we used the Lab colour space to calibrate the brightness of the images. In order to segment the LD muscle and extract the background we use an optimal linear combination of the RGB colour components with a threshold technique $(R > 80, G < 60, B < 120)$. Then, we applied morphological operations to correct the marbling regions as presented in Figure 1. We began with a close operation with a disk of 10 followed by a hole filling operation obtaining the region of interest.

After the segmentation of the region of interest area, the next step was to achieve the segmentation of the beef, in order to separate the

muscle without marbling from the total fat existing on the piece of meat. This operation was carried out applying morphological operations to segment the muscle from the intermuscular fat as shown in Figure 2.

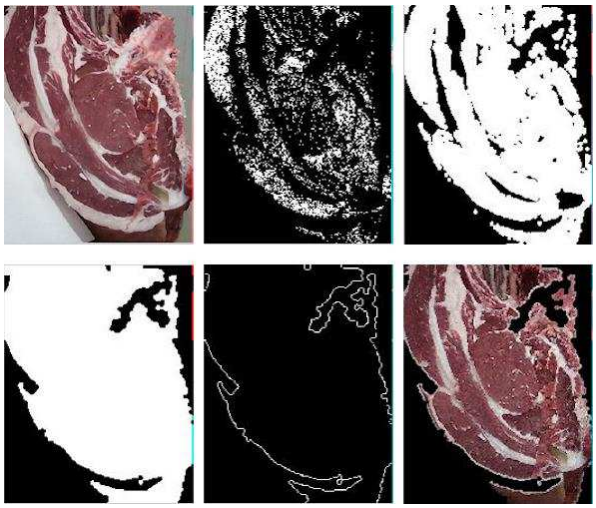


Figure 1: Extraction of the region of interest from the background.

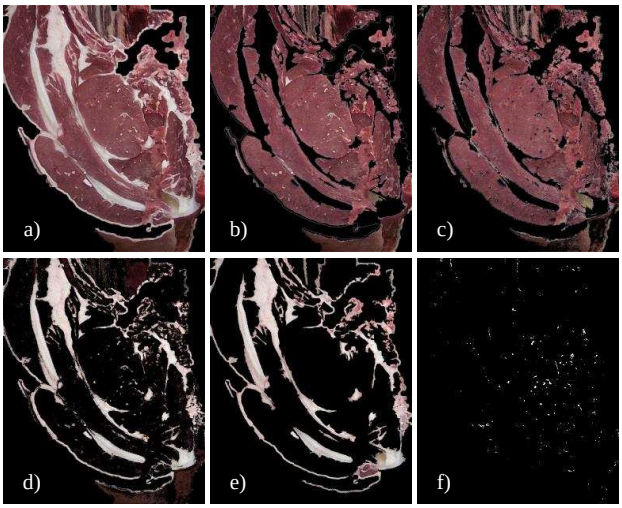


Figure 2: Extracted features: a) piece of beef; b) the muscle with marbling; c) the muscle without marbling; d) the total fat in a piece of beef; e) the intermuscular fat; f) intramuscular fat (marbling).

We obtained the total fat in the beef through a threshold technique based on the green component of RGB image (Figure 2.d). We removed marbling using morphological operations (Figure 2.e). In order to obtain the marbling Figure 2.f), we subtracted the image presented in Figure 2.e) from the beef image.

3 Results and Discussion

Figure 3 shows an example of the segmentation between the largest LD muscle area and its marbling.

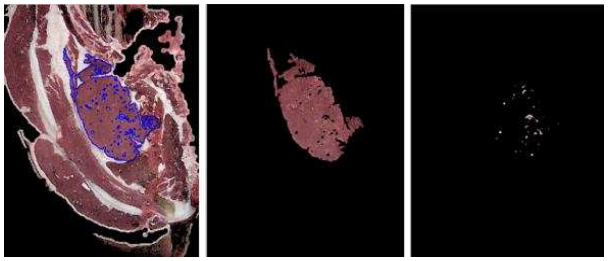


Figure 3: Left: original muscle image; middle: LD muscle image; right: marbling fat image.

To obtain the percentage of marbling we compared the marbling in each muscle with the total area of the muscle, as presented in Table 1.

Table 1: Percentage of marbling in the muscle of Figure 3.

Muscle LD	Muscle	Marbling
100%	98.12%	1.88%

We conducted our experiments using the proposed segmentation algorithm on the 83 sample images. Although the algorithm produce good results in most of the images in some of them it gives higher marbling percentage than it really exists in the beef. A reasonable explanation is that the moisture on the surface of beef rib-eye will cause reflection of light under a good lighting condition affecting to a degree the image processing which will be considered as marbling.

4 Conclusions

Image processing methods such as grey-scale transformation, contrast enlargement, morphological operations and binary segmentation, were used to process the image to effectively extract the rib-eye area of the cattle.

From the experimental results it has been confirmed that the proposed system enables high quality grading of beef marbling and robust region segmentation of the actual beef rib-eye image into lean and fat regions.

Acknowledgements

We thank to *Terra Fria Carnes* slaughterhouse in Bragança for their willingness in the acquisition of images.

References

[1] K. Chen and C. Qin. Segmentation of beef marbling based on vision threshold. *Computers and Electronics in Agriculture*, 62(2):223-230, 2008.

[2] W. Cheng, J.-H. Cheng, D.-W. Sun and H. Pu. Marbling analysis for evaluating meat quality: Methods and techniques. *Comprehensive Reviews in Food Science and Food Safety*, 14(5):523–535, 2015.

[3] P. Jackman, D.-W. Sun, C.-J. Du and P. Allen. Prediction of beef eating qualities from colour, marbling and wavelet surface texture features using homogenous carcass treatment, *Pattern Recognition*, 42(5):751–763, 2009.

[4] P. Jackman, D.-W. Sun, P. Allen. Prediction of beef palatability from colour, marbling and surface texture features of longissimus dorsi. *Journal of Food Engineering*, 96(1):151–165, 2010.

[5] B. Pang, X. Sun, K. Chen. A fast beef marbling segmentation algorithm based on image resampling. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(5):3894–3901, 2014.

[6] J. Subbiah, N. Ray, G. A. Kranzler, and T. Acton. Computer vision segmentation of the longissimus dorsi for beef quality grading. *Transactions of the ASAE*, 47(4):1261–1268, 2004.

[7] J. Tan. Meat quality evaluation by computer vision. *Journal of Food Enginnering*, 61(1):27–35, 2004.

Interactive/automated method to count bacterial colonies

João Ribeiro
a23422@alunos.ipb.pt
Ramiro Martins
rmartins@ipb.pt
Fernando Monteiro
monteiro@ipb.pt

Polytechnic Institute of Bragança, ESTiG,
Bragança, Portugal

Abstract

The number of colonies in a culture is counted to calculate the concentration of bacteria in the original broth; however, manual counting can be tedious, time-consuming and imprecise. Automation of colony counting has been of increasing interest for many decades, and these methods have been shown to be more consistent than manual counting. Significant limitations of many algorithms used in automated systems are their inability to recognize overlapping colonies as distinct and to count colonies on the plate boundary. This study proposes an interactive semi-automated counting system and a fully automated counting system using image processing methods which overcomes these problems. The proposed systems are capable to reduce the manpower and time required for counting colonies while taking account colonies both around the central area and boundary areas of a Petri dish. The results obtained are compared with other methods found in the literature.

1 Introduction

Bacterial culturing on solid agar plates (Petri dishes) is a fundamental process in microbiology, which is widespread for clinical laboratory exams, environmental control, food and beverage safety assessment.

The number of colonies in a culture is usually counted manually to calculate the concentration of bacteria based on the assumption that each colony has raised from one single bacterium (colony forming unit, CFU). However, this process is time-consuming (sometimes, the human who counts the colonies need to realize the procedure during many hours or even days), tedious (it is a monotonous procedure) and error prone (with the fatigue, the human being has more tendency to not do the right evaluation). The counting results obtained depend on the human conducting the count. This variability is one of the sources of error in the colony counting process that, along with methodological differences between different laboratories or even within a laboratory, can result in considerable fluctuations in results [1]. Due to this, for cultures with high density of colonies, manual counting mostly uses estimation methods, making an extrapolation from a small section of the Petri dish. Automating the detection, counting and analysis of CFU offers significant benefits to eliminate the risk of subjectivity, bias and human error, increasing speed and accuracy, and delivering unprecedented data archiving and retrieval capabilities.

Most automated counting systems perform adequately when the colonies are well spaced, large, circular in shape and with good contrast from the background. When these assumptions are violated, most automated colony analysis systems can rapidly lose reliability, accuracy and utility. These obstacles include the need to handle confluent growth or growth of colonies that touch or overlap other colonies; the identification of each colony as a unit in spite of differing shapes, sizes, textures, colours, light intensities; the exclusion of colonies around periphery of the plate reducing statistical accuracy.

To address the above problems, the goal of this study is to design and implement a cost-effective, software-centred system that accepts general digital camera images as its input, for detecting as well as enumerating bacterial colonies in a fully automatic manner. An interactive semi-automatic system is also proposed to overcome any error from fully automatic system. The proposed systems are capable to reduce the manpower and time required for counting colonies while producing correct colony counting.

2 Methods

In this study, we use a database of images available from [2] that contains 21 different images of bacterial cultures.

First, the original image is pre-processed in order to remove noise artefacts and to identify the dish area. The RGB image (colour images), shown in Figure 1, was converted into greyscale images. To remove the background, a mask was created from the original image and multiplied by the greyscale image.

Figure 1: Original image, retrieved from Chiang's database [2].

The second step of the method was to separate the central area of the Petri dish from the rim area. The greyscale image, in Figure 2(a), shows that the background of the central area is lighter than the surrounding of the rim area. Therefore, the central area, Figure 2(b), and the rim area, Figure 2(c), can be separated using thresholding processes.

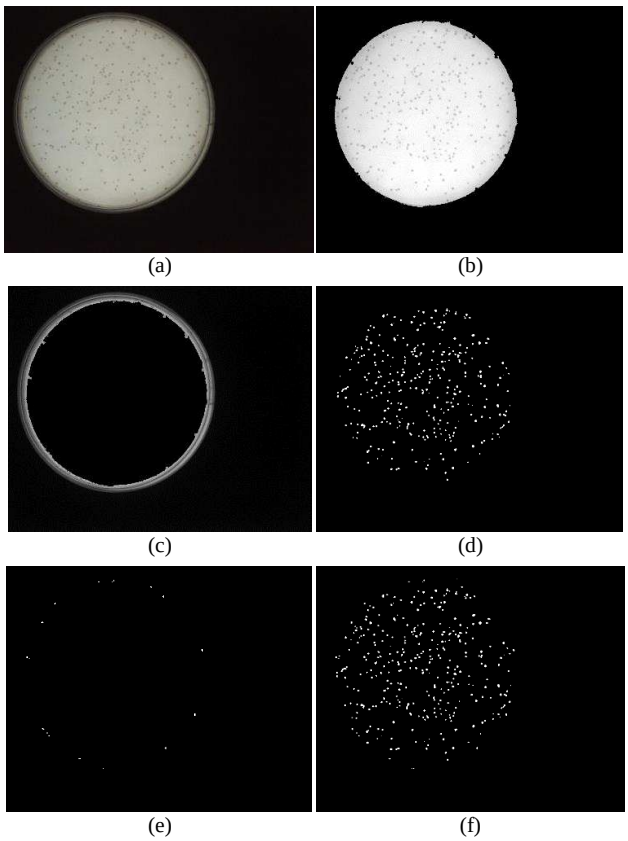


Figure 2: a) Original image; b) Central area; c) Image of the rim area; d) Colonies detected in central area; e) Colonies detected in the rim area; f) Sum of all detected colonies.

To obtain the Figure 2(d), a morphological close operation was applied, using a radius capable of cleaning the colonies from the central area. This image represents only the background from the Petri dish and

was subtracted by Figure 2(a). The final result in Figure 2(d) is an image where, supposedly, appears only the colonies.

To extract the colonies from the rim area was applied the bottom-hat transform. Then, knowing that the rim is longer and narrower than the colonies, some characteristics of the objects were calculated to obtain only colonies in the image, Figure 2(e). The Figure 2(f) represents the sum of the colonies of the central area and the colonies of the rim area.

In the next step, we obtained the area and eccentricity of each colony of the segmented image (Figure 2(f)). If an object has an area equal or smaller than the mean area of all the objects and an eccentricity lower 0.5 it is considered as an isolated colony, if not, it is considered as an overlapped colony.

The two methods (automatic and semi-automatic) use the same code to count the single colonies. It was applied the same method (to count colonies) used by Brugger [3], where a Bayes classifier is applied to count the final number of bacterial colonies. A Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem. Geometric properties, such us ratio between major and minor axis length of the group are used to verify the number of colonies contained on the image.

2.1 Automatic Method

To separate the clustered colonies it was used the watershed operation that computes a label matrix identifying the watershed regions of the input matrix, which can have any dimension. The elements obtained are integer values greater than or equal to 0. The elements labelled 0 do not belong to a unique watershed region. These are called watershed pixels. The elements labelled 1 belong to the first watershed region, the elements labelled 2 belong to the second watershed region, and so on.

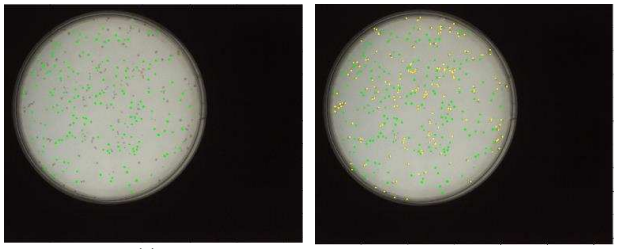
Although the watershed operation worked well to objects with few colonies, Figure 3(a), when the cluster has an aggregation of colonies, the watershed do not divide it uniformly, Figure 3(b). To overcome this issue, the mean area of all the unit colonies on Figure 2(f) was calculated. Thus, the larger areas, such as shown in Figure 3(b), were divided by the mean area obtained previously. These results were added to the count of watershed operation, and then added to the count of colonies.

(a) (b)

Figure 3: Results of the watershed transformation.

2.2 Semi-automatic Method

In the semi-automatic method, the algorithm counts all the separated colonies identifying them with the green colour, as shown in Figure 4 (a). The user marks each overlapped colony with a yellow point by clicking the mouse over the colony, or in the cluster of colonies, and decides how many colonies are in the clustering, as shown in Figure 4 (b). This number is then added to the number of isolated colonies, yielding the total number of colonies in the Petri dish.



(a) (b)

Figure 4: (a) Beginning of semi-automatic method, where the colonies on green are counted; (b) Result of semi-automatic counting, where the yellow points show the user's mouse clicks;

3 Discussion and Conclusion

To evaluate the performance of the two proposed methods, 21 images of colonies in Petri dishes were used. These images were counted automatically by the proposed systems and also manually by Biomedical Engineering students. The results obtained were compared with others 3 automatically counting systems (NICE [4], Clono Counter [5] and Chiang et al. method [2]).

The statistical results of the proposed systems were calculated after the automatic and semi-automatic counting. The statistical results of the systems NICE, Clono Counter and Chiang were obtained from [2]. The statistical results of precision, recall, F-measure and absolute percentage error (APE) comparing the methods are presented in Table 1.

Analysing the results of the automatic method, the value of precision is greater when comparing to the other three methods but the value of recall is the worst. The value of F-measure is greater than the value of Clono counter and worse than the other methods. Relatively to the APE, the automatic proposed system is 16,414% bigger than Chiang (3,37%) and NICE (7,2%) and smaller than Clono Counter (24,94%).

Table 1: Comparative results of precision, recall, F-measure and APE.

Method	Precision	Recall	F-measure	APE (%)
Automatic	0.99	0.85	0.91	16.41
Semi-automatic	0.99	0.99	0.99	0.79
Chiang	0.96	0,96	0,96	3.37
NICE	0.96	0,91	0,93	7.2
Clono Counter	0.79	0,95	0,85	24.94

Regarding the semi-automatic proposed system the results are better than all the existent methods. The statistical results obtained in this method (precision, recall, F-measure and APE) are close to perfection. It happens because even if the first phase of the semi-automatic method misses one colony, the interactive phase compensates this gap.

Comparing the aforementioned methods, the semi-automatic proposed method presents the best results.

References

[1] J. M. Bewes and N. S. McKenzie, "Automated cell colony counting and analysis using the circular Hough image transform algorithm (CHiTA)," *Physics in Medicine and Biology*, vol. 53, pp. 5991-6008, 2008.

[2] P. J. Chiang, Z. S. H. M. J. Tseng and C. H. Li, "Automated counting of bacterial colonies by image analysis," *Journal of Microbiological Methods*, vol. 108, pp. 74-82, 2015.

[3] S. D. Brugger, C. Baumberger, M. Jost, W. Jenni and U. Brugger, Automated counting of bacterial colony forming units on agar plates, *Plos one*, vol. 7, 2012.

[4] M. L. Clarke, R. L. Burton,A. N. Hill, M. Litorja, M. H. Nahm and J. Hwang, Low-cost, high-throughput, automated counting of bacterial colonies, *Cytometry A*, 77(8):790-797, 2010.

[5] M. Niyazi, I. Niyazi, and C. Belka, Counting colonies of clonogenic assays by using densitometric software. *Radiation Oncology*, 2(1), 1-3, 2007

Multi-modal Imaging Breast Cancer Detection: A Technological Review

João Teixeira[1]
jpft@inescporto.pt
Hélder Oliveira[1,2]
hfpo@inesctec.pt
Luís Teixeira[1,2]
luisft@fe.up.pt

[1] INESC TEC
Porto, PT
[2] Faculty of Engineering
University of Porto
Porto, PT

Abstract

Breast cancer is a disease that affects 1 in 10 women in Europe at some point in their lives. The 10-year survival rate after treatment is now exceeding 80% and Breast Conservative Treatment (BCT) has similar rates to traditional full breast removal (Mastectomy) while providing better aesthetic results. A fundamental task for this localized approach concerns the accurate detection and classification of potential cancerous tissues based on radiologic imaging. This paper describes various techniques and approaches towards this problem.

1 Introduction

Breast conservative surgery (BCT) is a breast cancer treatment option that tries to alleviate surgical results by combining Lumpectomy (local excision) with some form of radiation treatment, to avoid further cancer occurrence. This approach shares similar survival rates as Mastectomy (full breast removal). However, in BCT approximately 30% of women receive a suboptimal or poor aesthetic outcome.

An accurate identification of a tumorous mass is essential to predict the tissue to excise and/or transfer and thus, influences surgical planning. To this end, further study has been conducted on mass detection on radiological images. X-ray mammography, ultrasound and magnetic resonance (MRI) are the most widely accepted imaging techniques, each having their known advantages and limitations. These range from affordability, cancer under and overdetection rates to accurate mass location and ease of reproducibility. It is reported that certain types of cancer are only visible in certain imaging modalities. This means that there is an increased ability to detect a tumor if multiple imaging techniques are used, instead of regular mono-modal screening. However, due to several constraints, getting patients to undergo more than one exam is not always feasible.

Even though there have been relatively recent developments in breast imaging, namely Breast Tomosynthesis, Automated Whole Breast Ultrasound (AWBUS) and Ultrasound Computer Tomography (USCT) , this contribution will primarily focus on what is achievable using the conventional X-ray Mammography, MRI and Ultrasound techniques.

For obvious reasons, a single article is not able to review with detail all existing methods, however we hope that the references provided will cover the majority of significant work and steer researchers towards what has been done in the field. Our research goal is to develop breast models, combining medical imaging results according to exam availability, in order to detect more accurately breast tumors.

2 Single Modality Approaches

The majority of cancer detection approaches based on medical imaging search for suspicious regions image by image, in a 2D uncorrelated approach. In the case of X-ray Mammography, this is mainly due to the generic two-image acquisition protocol. For two distinct directions (mediolateral-oblique - MLO, and cranio-caudal - CC), the breast is highly compressed within two plates and an X-ray photography is taken. This imaging protocol seldom overlooks lesions since these can be concealed behind fibrous tissue. This, combined with the compressive nature of the exam, makes direct 3D reconstruction unfeasible. For MRI however, several images are taken across the torso and analyzed slice-by-slice. For Ultrasound imaging, depth images are taken across the breast towards the chest wall. Among the three, this is the least repeatable exam since, the combination of manual variable breast compression and sensor position, cannot be accurately reproduced.

2.1 X-ray Mammography

One of the initial tasks, usually performed on MLO view, Mammography lesion detection is the removal of the *Pectoral Muscle*. This task may be required because the muscle generally appears as a high intensity region on the mammogram and thus, might interfere with the lesion detection algorithms. In particular, the breast's parenchyma and lesions tend to also appear as high intensity structures.

The pectoral muscle appears at the top left or right corner and usually shares a variable length, fuzzy border with the breast portion of the image. The most basic impulse is to use Otsu's method to segment the intensities and obtain the connected components corresponding to the muscle. A more general approach consists of using this coarse thresholding, then detecting edges, growing the border contours and growing the pectoral muscle region. The path between two endpoints at the edges of the image can be optimized. Similarly, the shortest path is found considering the radial profile of the gradient, by treating the pectoral muscle as a round-like object. Despite these efforts, tumor cells can also appear in the pectoral muscle region and thus, many researchers overlook this task.

As in most image processing and classification problems, lesion detection approaches have more or less converged towards a generic method pipeline. This consists of four phases: pre-processing, feature extraction, feature set recombination and selection and, finally, the classification.

In the pre-processing stage, several filtering approaches can be use to enhance and transform the image, such as performing intensity expansion ($w(x,y)^{\alpha}$) on Wavelet space and converting back to image space.

Feature Extraction consists of acquiring pixel or window-wise image descriptors that characterize them. Currently, several types of image descriptors have been studied: *Intensity* based and *Statistical* features such as mean, standard deviation, skewness and kurtosis; *Texture* based (GLCM, LBF); *Morphological* and *Shape* descriptors; *Optical Density*; *Gabor* filter bank responses; Fractal and Ranklet features.

The total set of these features are far more than what is tractable with a limited computer in limited time. Thus, researchers usually pick a subset that presents the best change to accurately discern benign from malignant behavior. This can be done with *Recursive Feature Elimination* or *Best First Search*, trying to optimize some criteria for example, *Information Gain* or *Mutual Information*. *Feature Correlation* and *Linear Discriminant Analysis* also helps to decide which seem redundant. Furthermore, some techniques enable feature set reduction while also transforming it. *Principal Component Analysis* (PCA) and *Independent Component Analysis* (ICA) recombine the feature set, either by finding the best feature space representation or the features' non-Gaussian subcomponents, enabling a posterior, more relevant selection.

In the last stage, the remaining features are feed to classifiers such as *Decision Trees*, *Naïve Bayes*, *K-Nearest Neighbors*, *Artificial Neural Networks* (ANN), *Support Vector Machines* (SVM) in order to decide the result for that set of pixels. These models are generally first trained with ground truth data, for instance features extracted from regions with known benign or malignancy, providing examples to follow. In addition to the use a single classifier, combined classification schemes were developed namely, applying several similar or different classifiers and subjecting them to a weighed voting scheme (*Emsembles*: Bagging, Boosting).

There are also region-based approaches, that do not immediately try to find suspicious regions through clustering. For instance, Rouhi *et al*. [4] first tries to locate possible cancerous regions with Region Growing, with thresholds learned by a Neural Network, or segmented by a Cellular Neural Network, whose parameters are obtained by Genetic Algorithms. Only after this task, are the features obtained for the classifier to decide the benign or malignancy of the detected object.

There has been extensive work in X-ray, mainly because it is the old-

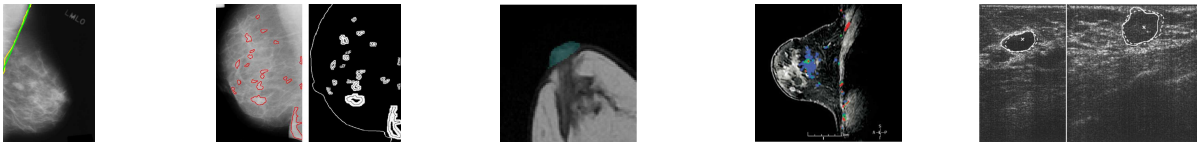


Figure 1: From left to Right: X-ray Pectoral Muscle Detection, X-Ray Mass Detection (two images), MRI Nipple Detection, DCE-MRI Contrast Analysis, Ultrasound Object Detection (two images).

est of the imaging methods, it is a relatively inexpensive exam and produces good enough results to be considered the lead screening test. Other advancements earlier than 2010, can be found in Oliver *et al*. [2] and the performances of more recent studies are compared in Tai *et al*. [5].

2.2 Magnetic Resonance Imaging

The modality with most morphological information is the MRI. With this imaging technique, a 3D representation can easily be formed from the 2D evenly spaced scans. As opposed to the view compressed mammography, MRI makes possible to discern the various breast structures. Most times several complete scans are performed using contrast material (Dynamic Contrast Enhanced MRI, or DCE-MRI) making way for simpler algorithmic approaches for lesion detection and classification.

Some approaches employ directly the pixels, or voxels (eg. intensity of neighborhood pixels as input of ANN), while others obtain descriptors that will be feed to the classifiers [1]. Many researchers report calculating time-intensity features when approaching DCE-MRI, such as variations of pixel/voxel-wise time-intensity curve. Others simply consider the local intensity increasing above a fixed or adaptive threshold, to be identified as suspicious, or even use the image/3D matrix obtained from subtracting the post-contrast to the pre-contrast information as input for feature extraction.

2.3 Ultrasound Imaging

Some effort has been devoted to pre-processing since Ultrasound generates very noisy images. The methods range from regular filtering to Wavelet techniques and Compounding (spacial) approaches.

For lesion segmentation, *Active Contours*, *Markov Random Fields* and *Region Growing* have been used. Classification approaches are also applied. Features of interest generally comprehend Texture and Morphological characteristics, along with Descriptor features such as *Echogenecity* and *Distortion Echogenic Halo*. In terms of classifiers, *Linear Discriminant Analysis* (LDA), *Logistic Regression*, *SVM* and *ANN* are frequently reported. Further details can be found in Rose and Allwin [3].

3 Multi-Modality Approaches

To improve mass detection accuracy other methods have been recently developed that combine information of X-ray Mammography and MRI. This combination generally comes in the form of registration, or fusion, where the almost uncompressed 3D breast model, obtained from an MRI, is virtually deformed until it presents the shape it would have when compressed for a given view of the Mammography exam. After that, the MRI information is projected onto the X-ray image and methods similar to those mentioned for the mono-modal can be coupled together.

Naturally, in order to produce the registration, one needs to define key regions in both modality spaces so that to properly align them. Concerning images from X-ray, the most common regions segmented consist of the nipple region, the breast limits (skin) and, in the case of the MLO view, also the pectoral muscle. MRI directly deals with 3D information and presents richer morphological detail, namely other organs, muscles, and bones. For breast cancer oriented studies, the segmentation structures of interest usually are similar to X-ray (3D equivalent), and may include the distinction of fatty and fibroglandular tissue.

Considering current 3D model technological options, some work has been developed on parametric model of the breast in which some models' physical properties are more transparent, though deformation manipulation becomes less intuitive, namely through *paraboloid primitives*; while others are less transparent but more intuitive concerning deformation manipulation, as using *Non-Uniform Rational Basis Spline* (NURBS).

There are also models derived directly from the variations of breast morphology. *PCA* has been used to compute generic external breast models which can be easily adapted to personalized information by altering the values of the major principal directions, or eigenvectors.

For physically accurate simulation, the *Finite Elements Method* (FEM) has been considered for emulating the exterior deformations of the skin and for modeling the breast interior, using different physical mechanics. However, regular FEM approaches have simulation time constraints in order to obtain their accuracy and need to be simulated off-line.

The majority of research on multi-modal, X-ray-MRI registration, concerning mass detection on the projection, has focused on one modeling technique, FEM. However, the approaches might differ somewhat. Some just use the pre-detected landmarks and try to find the transformation necessary to align the two modalities. Other researchers optimize this transformation and projection using fat and fibroglandular tissue shape. The MRI projection ("X-ray attenuation") makes possible to segment these keypoints in a similar fashion to the actual X-ray counterpart and the goodness of the comparison of these regions between both modalities becomes an optimization criteria. Another approach is to precisely measure the deformation that occurs within the breast during the X-ray exam compression. They measured the change in position of several landmarks inside a phantom of a breast that underwent a regular X-ray mammography exam. With this information, they can obtain the transformation model and can deform the MRI 3D model towards the X-ray view.

4 Conclusion and Future Work

According to the current literature, automatic lesion detection and classification has gone a long way since its inception and many research successes have been achieved. However, there is still some room for improvement concerning mass false detection. To this end, recent studies have focused on combining X-ray Mammography and MRI information.

Our goal is to continue to pursue this concept of multi-modality contribution and to continue to improve mass detection and classification results. We will study further non-cancerous regions of interest so to aid the registration task, registration methods less computationally expensive than FEM and also the possibility of including Ultrasound imaging information to the already accomplished X-ray-MRI space.

5 Acknowledgements

This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

[1] Monique D. Dorrius, Marijke C. Jansen van der Weide, Peter M. A. van Ooijen, Ruud M. Pijnappel, and Matthijs Oudkerk. Computer-aided detection in breast MRI: a systematic review and meta-analysis. *Eur Radiol*, 21(8):1600–1608, mar 2011.

[2] Arnau Oliver, Jordi Freixenet, Joan Martí, Elsa Pérez, Josep Pont, Erika R E Denton, and Reyer Zwiggelaar. A review of automatic mass detection and segmentation in mammographic images. *Med. Image Anal.*, 14(2):87–110, April 2010.

[3] R. Jemila Rose and S. Allwin. Computerized cancer detection and classification using ultrasound images: A survey. *International Journal of Engineering Research and Development*, 5(7):36–47, jan 2013.

[4] Rahimeh Rouhi, Mehdi Jafari, Shohreh Kasaei, and Peiman Keshavarzian. Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Syst. Appl.*, 42(3):990 – 1002, feb 2015.

[5] Shen-Chuan Tai, Zih-Siou Chen, and Wei-Ting Tsai. An Automatic Mass Detection System in Mammograms Based on Complex Texture Features. *IEEE J. Biomed. Health Inform.*, 18(2):618–627, mar 2014.

Outlier Detection in Ultrasound Using Blind Inpainting

Manya V. Afonso
mafonso@isr.ist.utl.pt
João M. R. Sanches
jmrs@isr.ist.utl.pt

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Lisbon, Portugal

Abstract

We propose a method for detection of outliers in Ultrasound images, by using blind image inpainting for the Rayleigh and multiplicative speckle noise model. Blind inpainting produces an estimate of the image as well as an estimate of the sampling mask which indicates which pixels are missing or corrupted. A logarithmic transformation is applied to convert the multiplication between the image, the binary mask, and the Rayleigh distributed speckle noise into an additive problem. The image and mask terms are then estimated iteratively with total variation regularization applied on the image, and ℓ_0 regularization on the mask term which imposes sparseness on the support set of the missing pixels. The resulting alternating minimization scheme simultaneously estimates the image and mask, in the same iterative process. Experimental results show that this method when applied on the RF image of the carotid artery without loss of pixels led to an outlier map which delineates clearly the lumen, and can therefore be applicable for segmentation.

1 Introduction

Ultrasound images are corrupted by speckle noise, which results from interference patterns. It has been shown that the observed ultrasound Radio Frquency (RF) image can be well approximated as an element-wise multiplication between the image pixels which represent the morphological structure of the organ being imaged, and the speckle field [9]. Despeckling methods seek to reduce the speckle noise and make the morphological structures more clearly visible. The spiky component of the speckle is modelled as outliers which do not fit in the statistical distribution assumed, for example Rayleigh. In [9], an outlier shrinkage step is applied on the log transformed image to eliminate these outliers. An adaptive window method to compute local statistics and to discard local extrema and replace them by average values based on the local statistics, on the B-mode image was proposed in [13].

Image reconstruction and inpainting methods for multiplicative noise generally require the pixel locations of the outliers to be known and therefore cannot be applied to estimate the pixel values at the locations of the outliers. In [3], a new method for blind image inpainting was proposed, which estimates the values of pixels which are missing or corrupted with impulse noise, when their locations are unknown. This method was also extended to non-additive and non-Gaussian noises. In this chapter, we review this method for the case of Rayleigh multiplicative speckle noise, and show that applying it on an Ultrasound image of the carotid artery can be useful for segmentation of the lumen.

2 Proposed Method

The image to be estimated has n pixels and is represented as a vector, say in lexicographic ordering, $\mathbf{x} \in \mathbb{R}^n$. Let $m < n$ be the number of observed or uncorrupted pixels. Observing a partial set of m pixels out of n can be represented as an element-wise multiplication of the image with a binary mask in which all but m pixels are zero. This process is represented as a multiplication of the vector $\mathbf{x}$ with a size $n \times n$ identity matrix $\mathbf{A}$ with the respective diagonal elements corresponding to the $(n-m)$ missing pixels set to zero.

For multiplicative noise, the mapping from $\mathbf{x}$ to the partially observed image $\mathbf{y}$ is given by $\mathbf{y} = \mathbf{A}(\mathbf{x}.\eta_S)$,where the speckle noise term η_S is Rayleigh or Gamma distributed, and the multiplication is element-wise. Multiplying a pixel whose value is 0 will always lead to the corresponding observed pixel being equal to 0 as well. Therefore, we interchange the order of the noisy observation and masking so that our observation $\mathbf{y}$ is the result of observing the masked image $\mathbf{Ax}$ under the noise model. The observation model changes to, $\mathbf{y} = (\mathbf{Ax}).\eta_S$.

For the Rayleigh multiplicative noise, the data fidelity term is,

$$J_r(\mathbf{y}, \mathbf{Ax}) = \sum_{i=1}^{n} \left(\frac{y_i^2}{2(a_i x_i)} + \log(a_i x_i) \right). \tag{1}$$

In (1), there appears a term with the product $(a_i x_i)$ and a term involving its logarithm. Therefore, we can work with the log transformed variables $\mathbf{u} = \log \mathbf{x}$, and $\mathbf{v} = \log \mathbf{a}$. Thus (1) changes to

$$J_r(\mathbf{y}, \mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} \left(\frac{y_i^2}{2} e^{-(u_i+v_i)} + u_i + v_i \right). \tag{2}$$

We now formulate our optimization problem, with TV regularization on $\mathbf{u}$ and ℓ_0 regularization on $\mathbf{v}$. The data fidelity term $J(.)$ is changed accordingly. The estimation problem is,

$$(\hat{\mathbf{u}}, \hat{\mathbf{v}}) = \arg\min_{\mathbf{u},\mathbf{v}} J(\mathbf{y}, \mathbf{u}, \mathbf{v}) + \frac{\lambda_1}{2} TV(\mathbf{u}) + \frac{\lambda_2}{2} \|\mathbf{v}\|_0, \tag{3}$$

where $TV(.)$ is the isotropic total variation [12] and $\|.\|_0$ is the ℓ_0 norm.

Since (2) involves the sum of a linear term and an exponential term, it is non-separable for $\mathbf{u}$ and $\mathbf{v}$. Therefore we need to use variable splitting [5] to be able to use the Augemented Lagrangian/Alternating Direction Method of Multipliers (AL/ADMM) to solve (3). We therefore introduce two auxiliary variables $\mathbf{z}$ and $\mathbf{w}$ to act as the arguments of the TV and ℓ_0 regularizer terms respectively, leading to the constrained problem,

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{w}} \quad & J(\mathbf{y}, \mathbf{u}, \mathbf{v}) + \tfrac{\lambda_1}{2} TV(\mathbf{z}) + \tfrac{\lambda_2}{2} \|\mathbf{w}\|_0 \\ \text{subject to} \quad & \mathbf{u} = \mathbf{z}, \mathbf{v} = \mathbf{w}. \end{aligned} \tag{4}$$

Using the augmented Lagrangian [8, 11], this problem can be shown to be equivalent to the minimization problem,

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v},\mathbf{z},\mathbf{w}} & J(\mathbf{y}, \mathbf{u}, \mathbf{v}) + \frac{\lambda_1}{2} TV(\mathbf{z}) + \frac{\lambda_2}{2} \|\mathbf{w}\|_0 + \\ & + \frac{\mu_1}{2} \|\mathbf{u} - \mathbf{z} - \mathbf{d_z}\|_2^2 + \frac{\mu_2}{2} \|\mathbf{v} - \mathbf{w} - \mathbf{d_w}\|_2^2, \end{aligned} \tag{5}$$

where $\mu_1, \mu_2 \geq 0$ are the penalty parameters, and $\mathbf{d_z}, \mathbf{d_w}$ are the so-called Bregman update vectors [7]. This problem is split into four problems at each iteration by gathering all the terms in each variable, and solving for each by keeping the others fixed. Thus, the AL algorithm iterates between minimizing the objective function in (5) with respect to $\mathbf{f}$ and $\mathbf{u}$, leading to a Gauss-Seidel process (for more details, see [1, 2, 6] and the references therein) which at iteration t is summarized as,

$$\mathbf{u}^{(t+1)} = \arg\min_{\mathbf{u}} J_r(\mathbf{y}, \mathbf{u}, \mathbf{v}^{(t)}) + \frac{\mu_1}{2} \|\mathbf{u} - \mathbf{z}^{(t)} - \mathbf{d}_z^{(t)}\|_2^2 \tag{6}$$

$$\mathbf{v}^{(t+1)} = \arg\min_{\mathbf{v}} J_r(\mathbf{y}, \mathbf{u}^{(t)}, \mathbf{v}) + \frac{\mu_2}{2} \|\mathbf{v} - \mathbf{w}^{(t)} - \mathbf{d}_w^{(t)}\|_2^2 \tag{7}$$

$$\mathbf{z}^{(t+1)} = \arg\min_{\mathbf{z}} \frac{\mu_1}{2} \|\mathbf{u}^{(t)} - \mathbf{z} - \mathbf{d}_z^{(t)}\|_2^2 + \frac{\lambda_1}{2} TV(\mathbf{z}) \tag{8}$$

$$\mathbf{w}^{(t+1)} = \arg\min_{\mathbf{w}} \frac{\mu_2}{2} \|\mathbf{v}^{(t)} - \mathbf{w} - \mathbf{d}_w^{(t)}\|_2^2 + \frac{\lambda_2}{2} \|\mathbf{w}\|_0 \tag{9}$$

$$\mathbf{d}_z^{(t+1)} = \mathbf{d}_z^{(t)} + \mathbf{z}^{(t+1)} - \mathbf{u}^{(t+1)},$$

$$\mathbf{d}_w^{(t+1)} = \mathbf{d}_w^{(t)} + \mathbf{w}^{(t+1)} - \mathbf{v}^{(t+1)}.$$

The ℓ_2-TV denoising problem (8) is solved using a few iterations of Chambolle's algorithm [4] and the ℓ_2-ℓ_0 regularized denoising problem from (9) is solved using the hard threshold. The problems involving $J_r(.)$, (6) and (7) can be solved approximately using a few iterations of Newton's method [10].

3 Experimental Results

All experiments were performed on MATLAB on an Ubuntu Linux based server with 64 GB of RAM. We applied our blind inpainting method on an RF ultrasound image to determine which pixels were statistically within the distribution and which ones had values that did not fit with it. The results obtained with the transversal and longitudinal images of the carotid artery are presented in figures 1 and 2, respectively.

In the estimated outlier maps in figures 1(c) and 2(c), the white pixels represent the mask pixels estimated incorrectly, *i.e.*, the result of a binary exclusive or (XOR) operation against the value one. Comparing these outlier maps with the respective RF and denoised images, it can be seen that the greatest concentration of outlier values is in the regions that correspond to the lumen. We also present the histograms for the speckle noise field computed by elementwise division of the observed image, by the denoised image $\hat{\eta} = \mathbf{y}/\hat{\mathbf{x}}$.

In figures 1(d) and 2(d), we present the histograms for the noise field over the pixel locations that are considered statistically valid, compared with the analytical probability density function for the Rayleigh distribution with parameter equal to one. For the transversal image, the Kullback-Leibler (KL) divergence between the histogram and the analytical distributions was found to be 0.074, with 13.92% of pixels labelled as outliers. Over the pixels marked as outliers, the KL divergence with respect to the analytical Rayleigh distribution increased to 2.49. For the longitudinal image, 21% of the pixels were marked as outliers, and the KL divergences with respect to the analytical Rayleigh PDF were 0.132 over the set of statistically valid pixels, and 3.67 over the outliers. The difference from the analytical curves can be seen in figures 1(e) and 2(e).

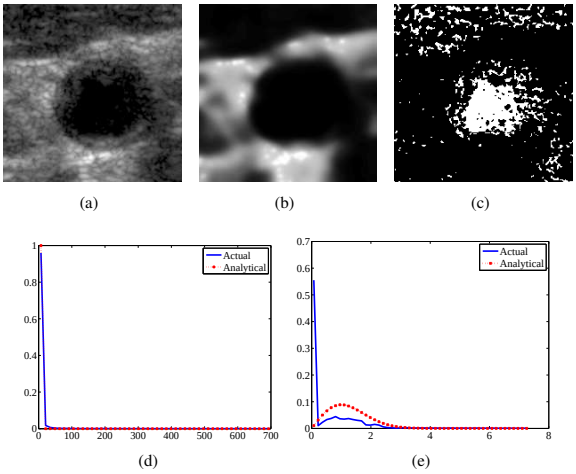


Figure 1: Blind inpainting applied on Ultrasound image to detect outliers: (a) transversal RF image of carotid artery, (b) denoised image; (c) positions of detected outliers; (d) histogram of speckle at valid pixels, compared with the Rayleigh distribution; (e) histogram of speckle at outlier positions, compared with the Rayleigh distribution.

4 Conclusions

We have applied blind image inpainting for the Rayleigh speckle noise model, to detect which pixels in Ultrasound images do not follow the expected statistical distribution. It was experimentally found for images of the carotid artery without loss of pixels, the pixels detected as outliers corresponded roughly to the lumen. Based on the results obtained with real ultrasound images, current and future research includes using the estimation of masks to help in obtaining optimal sampling patterns. A robust mathematical formulation for the segmentation problem is the subject of current and future research.

Acknowledgements

This work was supported by *Fundação para a Ciência e Tecnologia* (FCT), Portuguese Ministry of Science and Higher Education, through a Post-doctoral fellowship (contract no. SFRH/BPD/79011/2011) and FCT project (UID/EEA/50009/2013).

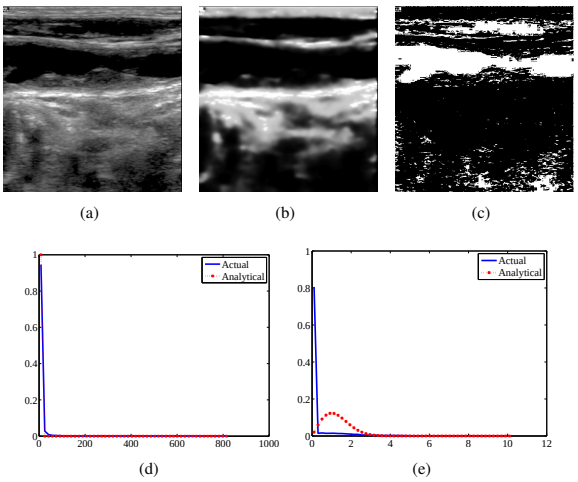


Figure 2: Blind inpainting applied on Ultrasound image to detect outliers: (a) longitudinal RF image of carotid artery, (b) denoised image; (c) positions of detected outliers; (d) histogram of speckle at valid pixels, compared with the Rayleigh distribution; (e) histogram of speckle at outlier positions, compared with the Rayleigh distribution.

References

[1] M. Afonso, J. Bioucas-Dias, and M. Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. on Im. Proc.*, 19(9):2345–2356, 2010.

[2] M. Afonso, J. Bioucas-Dias, and M. Figueiredo. An augmented lagrangian based method for the constrained formulation of imaging inverse problems. *IEEE Tr. Im. Proc.*, 20(3):681–695, 2011.

[3] M.V. Afonso and J.M. Raposo Sanches. Blind inpainting using ℓ_0 and total variation regularization. *Image Processing, IEEE Transactions on*, 24(7):2239–2253, 2015.

[4] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1): 89–97, 2004.

[5] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[6] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. Technical Report 09-31, CAM, UCLA, 2009.

[7] T. Goldstein and S. Osher. The split Bregman method for ℓ_1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[8] M. Hestenes. Multiplier and gradient methods. *Journal of optimization theory and applications*, 4(5):303–320, 1969.

[9] O.V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 53(1):64 –78, jan. 2006.

[10] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer verlag, second edition, 2006.

[11] M. Powell. *Optimization*, chapter " A method for nonlinear constraints in minimization problems"., pages 283–298. Academic Press, New York, 1969.

[12] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[13] P.C. Tay, S.T. Acton, and J.A. Hossack. Ultrasound despeckling using an adaptive window stochastic approach. In *Image Processing, 2006 IEEE International Conference on*, pages 2549–2552, Oct 2006.

Semi-supervised Approach for Liveness Detection in Fingerprint Recognition

Ana F. Sequeira
afps@inesctec.pt
Jaime S. Cardoso
jaime.cardoso@inesctec.pt

INESC TEC and Faculdade de Engenharia,
Universidade do Porto,
Porto, Portugal

Abstract

Fingerprint liveness detection methods have been developed as an attempt to overcome the vulnerability of fingerprint biometric systems to spoofing attacks. Traditional approaches have been quite optimistic about the behavior of the intruder assuming the use of a previously known material. This assumption has led to the use of supervised techniques to estimate the performance of the methods, using both live and spoof samples to train the predictive models and evaluate each type of fake samples individually. Additionally, the background was often included in the sample representation, completely distorting the decision process. Therefore, we consider that an automatic segmentation step should be performed to isolate the fingerprint from the background and truly decide on the liveness of the fingerprint and not on the characteristics of the background. Also, we argue that one cannot aim to model the fake samples completely since the material used by the intruder is unknown beforehand. We approach the design by modeling the distribution of the live samples and predicting as fake the samples very unlikely according to that model. Our experiments compare the performance of the supervised approaches with the semi-supervised ones that rely solely on the live samples. The results obtained differ from the ones obtained by the more standard approaches which reinforces our conviction that the results in the literature are misleadingly estimating the true vulnerability of the biometric system.

1 Introduction

Biometric recognition systems in general, and, in particular, fingerprint recognition systems (FRS), can be spoofed by presenting fake samples of the biometric trait to the system sensor. The fake samples can be acquired with or without user cooperation: an authorized user may help an hacker to create a clone of his fingerprint; or the fingerprint may be obtained from a glass or other surface. One possible and most successful method is to create a 3D fake model with the fingerprint stamped on it. This can be done by creating a mold that is then filled with a substance (silicon, gelatin, Play-Doh, wax, glue, plastic) and is used to create a thick or thin dummy that an intruder can use, see Fig. 1.

Figure 1: Finger Play-Doh mold and silicon model (from [3]).

Liveness detection methods can be categorized as hardware or software-based whether the detection is performed through additional hardware or by processing the obtained image. Software-based liveness detection methods can be divided in two categories based on whether they work with a single static 2D scan (static methods), or need 2D scans at different time points during the acquisition process that support the observation of the dynamic properties (dynamic methods). We confined this work to software-based approaches and within these to static features, in particular texture based statistics.

2 Methodological limitations of current research and proposed approach

One limitation of several of the existing procedures is the inclusion of the background in the liveness decision. As we are not interested in assessing the liveness of the background (which can be assumed to always be lifeless) but only of the fingerprint, it is therefore surprising to verify that fingerprint foreground segmentation although well established in the recognition works is often forgotten in the liveness detection works. To deal with this drawback we suggest the inclusion of an automatic fingerprint segmentation step before feature extraction for liveness detection. In the current experiments the method used is the one proposed by Ferreira et al. [1].

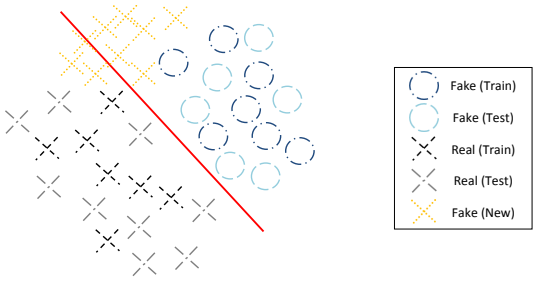


Figure 2: Illustrative Example. Black crosses and dark blue circles are fake and real samples in the training set. Light blue circles are real samples; gray crosses are fake samples from materials present in the training set; orange crosses are fake samples from a new material. The red curve represents the model learnt from the training samples (from [3]).

Another methodological limitation we pointed out is that models are designed and evaluated using fake samples of one type of material individually. Therefore, the systems are developed and tested under the assumption that the intruder will fabricate the fake fingerprint by employing one of the materials used for training, resulting in optimistically estimating the security level of the system. At the design time, the developer assumes to possess labeled data representative of the real and fake samples and therefore resorts to standard binary classifiers.

The binary classifiers adopted to make the decision between real and fake samples implicitly assume that the training samples are representative of the complete population, with the test data to which the system is applied coming from the same distribution as the training data. Although that may be a fair assumption for the real samples, it possibly is a crude model for fake samples created from a new material, see Figure 2. It may well happen that there is a mismatch between the distribution of the observations in training and testing data.

As new materials for fraudulent spoof attacks are going to continuously appear and become more and more sophisticated it is our conviction that the path to pursue is in the direction of using less and less information of spoofing materials and rely more strongly on the live samples. The study made presented (a) a realistic estimation of the performance in the presence of new materials used to fabricate the fake samples; (b) the use of decision models that rely only on the information from the real samples to detect the liveness.

The traditional approaches to liveness detection in general may be included in the category of *supervised detection* considering that they assume the availability of a training data set which has labeled instances for real and fake classes. Any unseen data instance is compared against the model to determine which class it belongs to. Having in mind that in real-world solutions the system is not "aware" of the kind of attack that might be performed, in this work we explore other approaches in which we do not assume knowledge about the fake samples used in a spoofing attack. In this study we use the real/live samples to train our model but not the fake/spoof ones. The techniques live one-class classification assume that the training data has labeled instances for only the live class. Since they do not require labels for the fake class, they are more widely applicable and do not overfit to the materials in the training set.

		Biometrika		Crossmatch		Italdata		Swipe	
		μ	σ	μ	σ	μ	σ	μ	σ
Traditional Approach	wLBP	2.00	1.40	**0.10**	0.01	3.17	1.09	9.53	2.02
	GLCM	7.73	5.38	5.32	2.97	**3.96**	1.98	7.63	0.71
Leave-one-out	wLBP	4.00	1.68	**2.78**	3.79	5.93	4.93	20.70	8.90
	GLCM	37.30	12.16	**35.96**	16.83	41.20	6.53	47.76	9.82
OCSVM	wLBP	**15.28**	4.58	20.13	10.53	24.48	5.05	42.68	6.07
	GLCM	**30.85**	14.53	31.73	10.16	37.78	3.81	40.76	5.23
GMM	wLBP	14.63	4.66	**1.44**	0.54	16.73	3.51	38.93	9.64
	GLCM	21.05	10.07	26.60	14.66	**12.95**	4.16	27.82	8.29

Table 1: Median Error Rate (MER) obtained with the different scenarios for the four datasets (MER in %).

3 Experimental Setup

3.1 Datasets

The methods were tested in the fingerprint datasets made available for the LivDet2013 competition [2]. The images of these datasets were acquired with four different sensors: Biometrika, Crossmatch, Italdata and Swipe and the fake samples were built using seven different materials: Body Double, Latex, Play-Doh, Wood Glue, Gelatin, Silicon and Modasil.

3.2 Feature Extraction Methods

The feature extraction was performed in the bounding box obtained from the segmented image. Two methods were applied: the Weighted Local Binary Patterns (wLBP) method [5] which combines Local Binary Patterns (LBP) with a Scale Invariant Feature Transform (SIFT); and the Gray Level Co-occurence Matrices (GLCM) method [4] which is based on GLCM matrices which characterize the relationship between neighboring pixels. For more details see [3].

3.3 Classification

For the classification task, different classification approaches were performed. We start by using a binary classifier and train and test within each dataset (containing samples made with only one type of fake material), then we introduce modifications to end up with a study where only the information about the live samples is used for training our models. The classifiers used were Support Vector Machines (SVM), one-class SVM (OCSVM) and a Mixture of Gaussians Model (GMM).

In the first two studies ("Traditional approach" and "Leave-one-out"), we used SVM with a polynomial kernel. A "grid-search" was performed on C and d for optimizing the parameters. In the one-class approach, we used a OCSVM and a GMM. Concerning the SVM, we used the RBF kernel and performed optimizing of the parameters by "grid-search" on γ and n parameters. The optimization of the parameters was performed by a nested validation procedure, therefore the evaluation is done in the test set which was not used for that purpose. Concerning the GMM, the number of components varied from 2^3 to 2^7 and the covariance matrix was not conditioned to be diagonal. For more details see [3].

4 Experimental Results and Discussion

In Table 1 are presented the results for the classification scenarios evaluated. In the first scenario, the best results obtained are 0.10% for wLBP and 3.96% for GLCM. Considering that, in practice, the system is not aware a priori of the kind of fake sample to be used in an attempt to fool the system the classification performed in method 1 is not very realistic. We do not intend to restrict our study to one material at a time, as traditional approaches do therefore the next scenario will be training will all materials except one that is left for testing. In this scenario we simulate that we know some possible attacks but we do not have knowledge about the one that is going to be perpetrated. In the "Leave-one-out" scenario the best results obtained are 2.78% for wLBP and 35.96% for GLCM. Regarding the one class approach: for the OCSVM the best results obtained are 15.28% for wLBP and 30.85% for GLCM; and for the GMM the best results obtained are 1.44% for wLBP and 12.95% for GLCM.

In an overall analysis, the results of the one-class approach are worse than the ones obtained by the first two scenarios. However, if we consider that in the first scenario we are training and testing with the same material, assuming to possess complete knowledge about the type of material to be used in a spoofing attack, then one important aspect is how much we are being optimistic when evaluating our models. A capable intruder will try to develop new materials for spoofing different from the known ones. So, even though the one-class approach results are worse globally we consider that this approach is more realistic than the one where the same material appears in the train and test phase assuming that the system developer will know all the spoofing materials. Other observation that can be made is that, for some sensors, the one-class approach is better than the "leave-one-out" which leads us to think that in some situations is better to use information only about the real samples and do not be mislead by information about fake samples which are significantly different from the fake samples which will be used to spoof.

5 CONCLUSIONS AND FUTURE WORK

In this study, different classification studies were performed in fingerprint liveness detection in order to broad the traditional approaches that use the knowledge about the fake/spoof samples for training the models. Therefore, the first innovative approach was to test with one material and train with the rest of the materials. As expected, this approach lead to worse results since we are using a complete unknown material in the test step. The approach which we consider the most worth following consisted on using only the information of the real samples when training our model and then test it with real and fake samples. In fact, what we are performing is a one-class classification characterizing the real samples and expecting our model to classify correctly as fake the spoof samples in the test set. Two different methods were used, a one-class SVM and a mixture of gaussians, and, in general, the best results were produced by this latter one. We note that the results of the one-class approach are worse than the other classification approaches, however our claim is that the first is more realistic. We consider more adequate to evaluate the robustness of a liveness method to unknown spoof attacks by not assuming complete or partial knowledge about the fake/spoof samples to be used by an intruder. We consider the results obtained in our experiments encouraging to pursue this approach in future works broadening the study to other databases and also other biometric traits. This approach, to the best of our knowledge, has not been yet fully studied and in our opinion can be further explored.

Acknowledgements

This work is financed by the FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

[1] Pedro M. Ferreira, Ana F. Sequeira, and Ana Rebelo. A fuzzy c-means algorithm for fingerprint segmentation. In Roberto Paredes, Jaime S. Cardoso, and Xose M. Pardo, editors, *Pattern Recognition and Image Analysis*, volume 9117 of *Lecture Notes in Computer Science*, pages 245–252. Springer International Publishing, 2015. ISBN 978-3-319-19389-2.

[2] Luca Ghiani, David Yambay, Valerio Mura, Simona Tocco, Gian Luca Marcialis, Fabio Roli, and Stephanie Schuckers. LivDet 2013 Fingerprint liveness detection competition. In *International Conference on Biometrics (ICB)*, pages 1–6, 4 - 7 June 2013.

[3] Ana F. Sequeira and Jaime S. Cardoso. Fingerprint liveness detection in the presence of capable intruders. *Sensors*, 15(6):14615, 2015. ISSN 1424-8220. doi: 10.3390/s150614615. URL `http://www.mdpi.com/1424-8220/15/6/14615`.

[4] Zhuoshi Wei, Xianchao Qiu, Zhenan Sun, and Tieniu Tan. Counterfeit iris detection based on texture analysis. In *19th International Conference on Pattern Recognition (ICPR 2008)*, pages 1–4. IEEE, 8 - 11 December 2008.

[5] Hui Zhang, Zhenan Sun, and Tieniu Tan. Contact lens detection based on weighted LBP. In *20th International Conference on Pattern Recognition (ICPR)*, pages 4279–4282. IEEE, 23 - 26 August 2010.

Experience-Based Planning Domains: An Approach to Robot Task Learning*

Vahid Mokhtari[1]
mokhtari.vahid@ua.pt
Luís Seabra Lopes[12]
lsl@ua.pt
Armando J. Pinho[12]
ap@ua.pt

[1] IEETA, University of Aveiro, Aveiro, Portugal
[2] DETI, University of Aveiro, Aveiro, Portugal

Abstract

This paper presents a formalization of experience-based planning domains, where experiences are collected, task knowledge is learned from experiences and plans are generated based on the learned knowledge. A learning framework is presented for a robot to: (*i*) obtain robot activity experiences from its own performance in a dynamic environment; (*ii*) conceptualize each experience to a unit of task knowledge called *activity schema*; and (*iii*) exploit the learned activity schemata to make plans in similar situations. Experiences are episodic descriptions of plan-based robot activities including environment perceptions, sequences of applied actions and achieved tasks. The conceptualization approach combines deductive generalization, abstraction, goal inference and feature extraction. A high-level task planner is proposed to find a solution for a similar task by following an activity schema. The proposed approach is illustrated in a restaurant environment where a service robot learns how to carry out complex tasks.

1 Introduction

Autonomous intelligent robots are required to *interact* with their environment, *acquire knowledge* to adapt to dynamic and changing environments, and *act deliberately* in order to achieve their mission. Acting deliberately means performing actions that are selected based on reasoning motivated by some intended objectives. Deliberation and learning are essential to endow autonomous intelligent robots with adaptable and robust capabilities, as well as reducing the deployment cost [1, 4].

The work described in this paper is based on the general idea of *learning from experience*, where a human user instructs and teaches a robot how to perform a complex task. It is an extension of our earlier works [2, 3], in which we proposed a learning system to obtain robot activity experiences, conceptualize experiences and make plans based on the learned concepts. Figure 1 provides an overview of the proposed learning and planning system. This system allows for: (*i*) verbally instructing a robot how to carry out complex tasks using a set of predefined primitive behaviors and previously learned concepts; (*ii*) extracting plan-based robot activity experiences from working memory contents; (*iii*) conceptualizing robot activity experiences and constructing activity schemata; and (*iv*) to close the loop, exploiting activity schemata by a planner to make plans for similar situations.

2 Formalizing experience-based planning domains

A formal definition of experience-based planning domains that evolve through learning from experiences is presented.

Definition 1 *An experience-based task planning domain is a tuple,*

$$\mathcal{D} = (\mathcal{L}, \Sigma, \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{E}, \mathcal{M}), \tag{1}$$

where $\mathcal{L}$ is a first-order logic language, Σ is a set of ground atoms of $\mathcal{L}$ that are always true (i.e. static world information), $\mathcal{S}$ is a set of states in which every state $s \in \mathcal{S}$ is a set of ground atoms of $\mathcal{L}$ which may become false (i.e. transient world information), $\mathcal{A}$ is a set of abstract operators, $\mathcal{O}$ is a set of planning operators, $\mathcal{E}$ is a set of plan-based robot activity experiences stored in the experience memory, and $\mathcal{M}$ is a set of methods in the form of activity schemata stored in the concept memory.

*This work was funded by the EC 7th FP theme FP7-ICT-2011-7, grant agreement no. 287752 (project RACE - Robustness by Autonomous Competence Enhancement), and by FCT scholarship SFRH/BD/94184/2013 and FCT project UID/CEC/00127/2013.

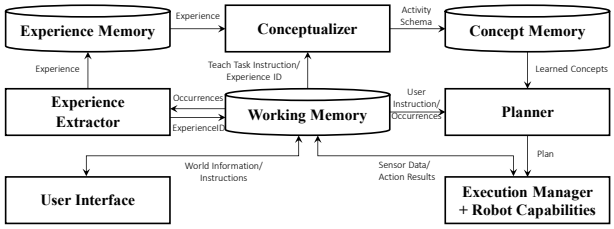


Figure 1: An overview of the learning and planning system.

Definition 2 *An abstract operator $a \in \mathcal{A}$ is a class of planning operators, which is described by a head*[1]*,*

$$a = (h), \tag{2}$$

where h is the head of a.

Definition 3 *A planning operator $o \in \mathcal{O}$ is a tuple,*

$$o = (h, a, S, P, E), \tag{3}$$

where h is the operator's head, a is an abstract operator which is the superclass of o, S is the static world information, and P and E are respectively the preconditions and effects of o.

Definition 4 *A plan-based robot activity experience $e \in \mathcal{E}$ is a triple,*

$$e = (t, \pi, K), \tag{4}$$

where t is the head of a task, taught by a human user to a robot, π is a sequence of applied actions to achieve t, and K is a set of propositions in the form of first-order binary predicates with qualitative timestamps [3].

Definition 5 *An activity schema $m \in \mathcal{M}$ is a triple,*

$$m = (h, G, \Omega), \tag{5}$$

where h is the head of the target task, G is a set of generalized goal propositions *inferred from an experience, and Ω is a sequence of abstract operators enriched with features.*

Definition 6 *An enriched abstract operator $\omega \in \Omega$ is a pair,*

$$\omega = (a, F), \tag{6}$$

where $a \in \mathcal{A}$ is the head of an abstract operator, and F is a set of features, in the form of reified fluents, that documents the arguments of a [2].

Definition 7 *A task planning problem is a triple,*

$$\mathcal{P} = (\sigma, s_0, t), \tag{7}$$

where $\sigma \subseteq \Sigma$ is the static world information, $s_0 \in \mathcal{S}$ is the initial state, and t is the task to be planned (e.g. `serve_coffee(guest1)`).

Definition 8 *A plan solution π to a task planning problem is a sequence,*

$$\pi = \langle a_1, ..., a_{k>0} \rangle, \tag{8}$$

where every a_i is an action. Plan π is generated for task t iff there is a learned activity schema $m \in \mathcal{M}$ corresponding to task t in which π can be derived from m.

[1] In this work, a *head* of an operator, task, etc is in the form $n(x_1, ..., x_{k\geq 0})$, in which n is the name, and $x_1, ..., x_k$ are the arguments, e.g. `move(?from ?to)`.

Algorithm 1 Conceptualizing a robot activity experience

Input:
a planning domain $\mathcal{D} = (\mathcal{L}, \Sigma, \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{E}, \mathcal{M})$ ▷ Def. 1
a robot activity experience $e = (t, \pi, K)$ ▷ Def. 4
Output:
an activity schema m ▷ Def. 5
procedure CONCEPTUALIZE($\mathcal{D}, e$)
  Build an explanation of e using $\mathcal{D}$ ▷ as in EBG
  Generalize e using a goal regression algorithm
  Apply an abstraction hierarchy on the generalized operators in π
  Infer goal propositions g from K and add it to m
  for each generalized abstract operator $a \in \pi$ **do**
    $a_{rich} \leftarrow (a, F(a,t,K))$ ▷ features extraction [2]
    $m \leftarrow$ add a_{rich}
  return m

```
1. [ serve_coffee(mug1,guest1),  %% Task
2.  [....],                      %% KeyProps (omitted)
3.  [                            %% SeqOp
4.   tuck_arms(aunp,aunp,aunp1,aunp0,atp,atp,atp7,atp13),
5.   move_base(at0,fatr1,at7,pmaec1),
6.   move_torso(torso1,tdp,tdp0,tup,tup2),
7.   tuck_arms(atp,atp,atp7,atp13,atp,atp,atp1001,atp17),
8.   move_arm_to_side(rightarm1,atp,atp17,asp19),
9.   move_base_blind_to_ma(at7,pmaec1,at9,maec1),
10   pick_up_object(mug1,rightarm1,on4,at9,maec1,paerc1,
                    rg1,h1,tup2,asp19,asp27),
11.  move_base_blind_to_pma(at9,maec1,at11,pmaec1),
12.  move_torso(torso1,tup,tup2,tdp,tdp4),
13.  move_arms_to_carryposture(atp,asp,atp1001,asp27,acp31,acp33),
14.  move_base(at11,pmaec1,at13,pmast1),
15.  move_torso(torso1,tdp,tdp4,tup,tup6),
16.  move_arm_to_side(rightarm1,acp,acp33,asp35),
17.  move_base_blind_to_ma(at13,pmast1,at16,mast1),
18.  place_object(mug1,rightarm1,at16,mast1,on3,pawrt1,rg1,
                  h1,tup6,asp35,asp42),
19.  move_base_blind_to_pma(at16,mast1,at19,pmast1) ]]
```

Listing 1: A plan-based robot activity experience for serving a coffee to a guest. It contains 16 primitive actions and 122 key propositions.

3 Conceptualizing experiences

An activity schema is a unit of task knowledge obtained from a plan-based robot activity experience. The term *conceptualization* in this project refers to the process of forming a new concept by assigning a description and a name to a real-world subset or pattern. The proposed conceptualization approach is comprised of different modules including deductive generalization, abstraction, goal inference and feature extraction [2, 5]. The conceptualizer takes as input a planning domain and a plan-based robot activity experience and constructs a relevant task knowledge as an activity schema. The general description of the conceptualizer procedure is represented in Algorithm 1. The learned activity schemata are eventually recorded into the concept memory which are later used for problem solving.

4 Planning based on activity schemata

An adapted A^* heuristic search planner, *Schema-Based Planner* (*SBP*), is proposed to find a solution for a given problem. The SBP planner takes as input a planning domain $\mathcal{D} = (\mathcal{L}, \Sigma, \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{E}, \mathcal{M})$ and a task planning problem $\mathcal{P} = (\sigma, s_0, t)$. If there is an activity schema, $m \in \mathcal{M}$, for task t, then the planner generates a plan for t; otherwise it fails. The SBP searches forward by selecting applicable actions belonging to the class of abstract operator ω in m, which verify a greater number of features. A feature is a fluent or a set of fluents, which relate an argument of an operator to an argument of the taught task in an experience. Features capture normative principles and support similarity-based planning. The description of the conceptualization and planning approaches was presented in more detail in [2].

5 Experimental results

The learning approach was demonstrated and evaluated in two different demonstrations, "serve a coffee" and "clear a table". To learn an activity schema, a plan-based robot activity experience is extracted from the memory through a human-robot interaction [3]. Listing 1 shows a plan-based robot activity experience for "serve a coffee" scenario. The conceptualization is carried out immediately after a plan-based robot activity experience is generated. Listing 2 shows a learned activity schema after generalization, abstraction, goal inference and feature extraction. An online video of the conceptualization process for "serve a coffee" scenario in real *PR2* robot is available in `http://youtu.be/lF33FmL1mjw`.

```
1. [ serve_coffee(MUG,GUEST),     %% Task
2.                                %% Goal
3.  [on(MUG,PAWRT),hasplacingarearight(SAWT,PAWRT),at(GUEST,SAWT)]
4.  [                             %% Enriched SeqOp
5.   [move(PATR1,PMAEC),[]],
6.   [move(PMAEC,MAEC),
7.     [at_start(on(MUG,PAERC)),
8.      throughout(hasmanipulationarea(PAERC,MAEC))]],
9.   [pick_up(MUG,PAERC),
10.    [at_start(on(MUG,PAERC)),
11.     at_goal(on(MUG,PAWRT))]],
12.  [move(MAEC,PMAEC),[]],
13.  [move(PMAEC,PMAST),
14.    [throughout(haspremanipulationarea(TABLE,PMAST)),
15.     throughout(hassittingarea(TABLE,SAWT))]],
16.  [move(PMAST,MAST),
17.    [throughout(hasmanipulationarea(PAWRT,MAST))]],
18.  [place(MUG,PAWRT),
19.    [at_goal(hasplacingarearight(SAWT,PAWRT)),
20.     at_goal(at(GUEST,SAWT))]],
21.  [move(MAST,PMAST),
22.    [throughout(haspremanipulationarea(TABLE,PMAST)),
23.     throughout(hassittingarea(TABLE,SAWT))]] ]]
```

Listing 2: A learned activity schema for serving a coffee to a guest containing 8 abstract operators. Line 3 shows a generalized goal which is instantiated based on a given task problem and should be achieved during planning. Some abstract operators are enriched with features. During planning, these features give preferences to find the best match action to the one in the activity schema. For instance, features at lines 19 and 20 help the SBP planner to among all applicable actions belonging to the class of *place*, prefer an action that closely matches these features in a given task problem.

Table 1: Planner performance metrics in two scenarios.

Activity schema*	srv(G,C)	srv(M,G)	srv(M,G,C)	clr(T)	clr(T,C)	clr(M,T,C)
Plan length	17	17	17	16	16	16
Nodes expanded	920	330	192	545	182	182
Search tree size	3706	1410	854	2247	765	765
Penetrance (%)	1.84	5.15	8.85	2.93	8.79	8.79
Average branching factor	4.02	4.26	4.44	4.12	4.19	4.19
Effective branching factor	1.62	1.53	1.48	1.61	1.51	1.51
Precision	0.94	1.00	1.00	0.81	1.00	1.00

* *srv* and *clr* stand for *serve_coffee* and *clear_table* tasks, and M, G, T and C stand for MUG, GUEST, TABLE and COUNTER.

The measures of *penetrance*, *effective branching factor* and *precision* are used to evaluate the performance of the SBP planner over the learned activity schemata [2]. Different activity schemata of "serve a coffee" and "clear a table" with different task arguments have been conceptualized. We performed one experiment for every learned activity schema (Table 1). For every demonstration, the penetrance ratio increased and the effective branching factor decreased with respect to the number of arguments of activity schemata. An online video to show the performance of the proposed SBP planner for "serve a coffee" scenario in real *PR2* robot is available in `http://youtu.be/JXuq4p_QCJk`.

6 Conclusion

This paper proposed: (*i*) a formalization of experience-based planning domains; (*ii*) a unified framework for solving problems in these domains; and (*iii*) an approach for conceptualizing experiences. The terms *experience*, *activity schema* and *schema-based planning* are tightly involved in the proposed planning domain. This system provides a learning and deliberation approach for intelligent robotics.

References

[1] F. Ingrand and M. Ghallab. Deliberation for autonomous robots: A survey. *Artificial Intelligence*, 2015. ISSN 0004-3702.

[2] V. Mokhtari, L. Seabra Lopes, A.J. Pinho, and G.H. Lim. Planning with activity schemata: Closing the loop in experience-based planning. In *Autonomous Robot Systems and Competitions (ICARSC), 2015 IEEE International Conference on*, pages 9–14, April 2015.

[3] Vahid Mokhtari, GiHyun Lim, Luís Seabra Lopes, and ArmandoJ. Pinho. Gathering and conceptualizing plan-based robot activity experiences. In *Intelligent Autonomous Systems 13*, volume 302 of *Advances in Intelligent Systems and Computing*, pages 993–1005. Springer International Publishing, 2016. ISBN 978-3-319-08337-7.

[4] L. Seabra Lopes and J.H. Connell. Semisentient robots: routes to integrated intelligence. *Intelligent Systems, IEEE*, 16(5):10–14, Sep 2001. ISSN 1541-1672.

[5] Luís Seabra Lopes. Failure recovery planning for robotized assembly based on learned semantic structures. In *IFAC Workshop on Intelligent Assembly and Disassembly (IAD'2007)*, pages 65–70, 2007.

Low Cost Vision-Guided Parking System

David Saraiva
david.vsaraiva@gmail.com
João M.F. Rodrigues
http://w3.ualg.pt/~jrodrig/

Instituto Superior de Engenharia, LARSyS,
University of the Algarve
Faro, Portugal

Abstract

Intelligent parking systems can reduce traffic, making it easier for users to find free parking space, which saves them time and money. However, these systems, often, require expensive installations. In this article, a proposal for a low cost computer vision framework based on Raspberry Pi 2 and an HD webcam for the detection of free parking spaces is presented. The vision algorithm is based on a simple edge detection, and a dynamic threshold within each parking space. The system was tested in a real environment with different weather conditions and partial occlusions. The algorithm runs in real time and has a 99.6% of success rate.

1 Introduction

Intelligent parking systems can reduce traffic more than 30% in the implemented areas, the proposed solution started from the need of the staff at the University of West Bohemia, Pilsen, Czech Republic for a more efficient parking system due to the time spent looking for a parking space. This system should be simple to maintain and as low cost as possible.

Intelligent parking systems is a widely studied subject [3, 5, 6, 7]. There are a set of algorithms [6, 7] that use characteristics of the car (hood, windows, headlights, bumpers, etc.), but they are not well suitable for cases where there are partial occlusions. Others use background subtraction, probably those are the most widely used in video-vigilance applications [2, 4]. However, there are several challenges to overcome, such as: gradual and sudden variations in lighting, shadows, e.g., the effect of the wind in the trees, and the speed of the objects. These challenges bring implications on the update of the background ratio and the learning rate. The intended solution should be simple to maintain, easy to install and low cost. A computer vision system was chosen, based on a HD webcam and a low cost processing unit (Raspberry Pi 2).

The system has 2 steps: (a) initialization, where the parking spaces were selected and (b) parking space validation, which is based on edge detection, the extraction of specific regions of each parking space, and the computation of the dynamic threshold ratio between the edge pixels that represent a car and the total pixels per parking space. The main contribution is a simple, low cost, easy to maintain system, prepared to work in different weather conditions (sun, rain, clouds) and capable of coping with partial occlusions of the cars.

2 Vision-Guided Parking System

The parking lot of the University of West Bohemia is quite big, it was divided in 9 sectors, where each sector has a camera (HD Webcam, 720p with 30fps) and a Raspberry Pi, see Fig. 1 top row, where each color represents the field-of-view of the camera, sector, and the interception between 2 cameras by a different color. The total cost of the system per sector is around 70 euros. The cameras are to be placed on the top of the building. In this article, sector H was used for the explanations and illustrations. There are two different park regions: the *bottom*, which has the parking spaces transverse to the camera and with the possibility of partial occlusions between the cars and the *upper*, which has the parking spaces in a longitudinal way but with "rotation". As mentioned, the system is divided in 2 modules: (i) initialization and (ii) parking space validation.

System Initialization (i), is done only once in the setup of the system. Consider $I_t(x,y)$ a frame acquired at each Δt = 1s (this value can be changed in the database (DB) of the application). With the initial acquired park frame I_0, the first step is to set manually, using an application provided to the park administrator, a sets of points that limit the parking spaces: (a) four points ($P_{\{b1,\ldots,b4\}}$) that define the corners of the bottom region of the park, see Fig. 1, the orange dots in the 2nd row. Later, those points will be used to get a ratio between the distance to the camera and the size of the car. (b) Four points ($P_{\{u1,\ldots,u4\}}$) that define the upper region of the park. Later, those will be used to do a perspective transformation of the upper part of the park, see Fig. 1, the red dots in the 2nd row, with the

Figure 1: System initialization, from top to bottom: Park sectors, image (I) from section H, RoI (R_i) of the upper and bottom regions of the part in green (color dots, important points; see text).

perspective transformation in the 3rd row. (c) A group (N_{pb}) of 4 points ($p_{\{1,\ldots,4\}}$), that mark the parking spaces in the bottom region of the park, $PP_{\{i,p\}}$, with $i = \{1,\ldots,N_{pb}\}$. The coordinates of each parking space are selected from the bottom boundary line of the parking space, to where the top of the car/truck is expected (and not to the 2nd boundary parking line). This way, the "center" of the parking space will go up, and it is expected not be highly covered by the bottom car, when the partial occlusion happens (the angle from the camera to the parking lot has to be such that allows this assumption). See Fig. 1, bottom row, the dot color dots. As can bee seen, it is not necessary a great precision in marking the points, only a draft prevision where the car/truck could appear. (d) The same as in c), but now, for the upper region defined in b), after applying the perspective transform. Defined by a group (N_{pu}) of 4 points $PP_{\{i,p\}}$, now with $i = \{N_{pb},\ldots,N_{pb}+N_{pu}\}$. The total number of spaces in each section is $N_{ps_s} = N_{pb_s} + N_{pu_s}$, with s the section letter/number. For section H, $N_{pb_H} = 12$, $N_{pu_H} = 8$ and $N_{ps_H} = 20$. For the total of parking spaces a sum of all sections $N_{ps} = \sum_{s=\{A,\ldots,I\}} N_{ps_s}$. Note, that s is only presented in this paragraph and removed from all the remaining paragraphs to simplify the notation. The remaining paragraphs are always referring to a single

Figure 2: Top, edge map (I_c), bottom and example of two R_i, left for bottom, right for upper.

section, and illustrations shows section H.

The *parking space validation* (ii) consists in 5 steps: (a) The first consists in applying the Canny edge detector to each I_t frame, returning I_c (the thresholds used follow the recommendation of [1]). The goal is to retrieve only the cars edges, however if edges from puddles, noise and tree shadows appear, they have no significant impact on the presented system. See Fig. 2 top row.

(b) For each PP_i a R_i image corresponding to a region of interest (RoI) was computed based from I_c, with $i = \{1,..,N_{ps}\}$. Each R_i allows the determination of a car presence. There are two methods to calculate these RoIs, depending on the bottom or upper region of the park. Considering the $PP_{\{i,p\}}$ points: (b.1) the bottom region RoIs, the car near the camera has a larger size than the cars far away from it. For computing the RoI (green rectangles in Fig. 1 bottom row) a percentage of 20% of the height of each parking space was used, and for the width, 3/4 of the parking space width, to remove any vegetation, etc. One example can be seen in Fig. 2, bottom row left. (b.2) The upper part, after the perspective transformation all parking spaces have almost the same size (rectangles in Fig. 1, 3rd row). The ROIs were obtained using 3/4 of the height and 40% of the width, see example in Fig. 2, bottom row right.

(c) The next step computes the ratio (r_i) between the white pixels (noise or car edges) by the total size of pixels in each RoI, i.e., $r_i = \sum_{R_i=255} R_i(x,y) / \sum R_i(x,y)$. In the presence of a car the count of white pixels is expected to be high. Two detection thresholds were computed: (c.1) The upper and bottom detection thresholds were obtained empirically by analyzing the correspondent RoIs (upper and bottom). (c.1) For the upper region of the park, a ratio threshold of $r_{tu} = 0.2$ was used. (c.2) For the bottom region, a static threshold is not suitable due to the impact of the perspective of the camera to the car, i.e., the effect on the size of the car and the "thickness" of the edges. In this case a linear equation was calculated to express the ratio between the distance and the size of the car, using the points $P_{\{b1,...,b4\}}$ (see Fig. 1, 2nd row), i.e., $b = (((P_{b3x} - P_{b2x}) - (P_{b4x} - P_{b1x})) \times (a + (P_{b4x} - P_{b1x})))/(P_{b3y} - P_{b4y})$. In this equation, the b represents the width of the park, for some height a, which gives the location of the parking space in the image I_t. Consider now the bigger car possible the "unity", where the width is given by $P_{b1x} - P_{b4x}$, the ratio that compares the size of the car with the height, in respect of the width is: $1.0/r_c = (P_{b4x} - P_{b1x})/b \Leftrightarrow r_c = b/(P_{b4x} - P_{b1x})$. Relating the upper and bottom regions a threshold $r_{tb} = r_{tu}/(4 \times r_c)$ is computed. (d) Finally, the upper (r_{tu_i}) and bottom (r_{tb_i}) thresholds depending on R_i are compared with r_i, if they are lower it means that a car is not present, see Fig. 3. If a parking space, presents as "occupied" for 5 seconds, the place is given as "occupied" and marked with red dot. In case of getting 5× the place as "free", the place is marked with green dots and given as free. In the case of state variation in the last 5 seconds period, the place is marked with yellow dots and given as state of "transition".

3 Experimental Result

For the tests, frames were acquired between sunrise (6 AM) and (8 PM), with $\Delta t = 1$s, during 8 days (except the weekend), from 15 to 22 of June of 2015, in the parking lot of the University of West Bohemia, with several weather conditions (clouds, short periods of sun, and rain). Most of the

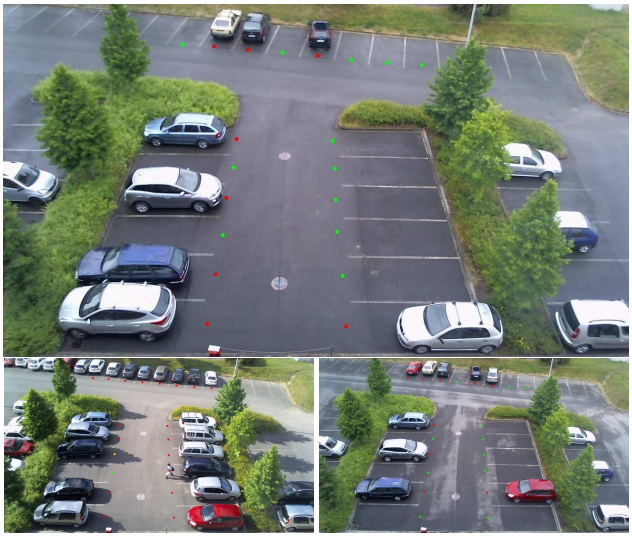


Figure 3: Results. Red dots, space *occupied*, green *not occupied* and yellow *occupied*, but in the previous frame returned *not occupied*.

times (minutes) there are no changes in a park (car parking or leaving). For that reason, a set of 82 images was chosen, around 10 per day, which corresponds to 1640 parking spaces (82×20, section H). For the tests, a DB was created by hand with the ground truth of the park. The system was tested and a success rate of 99.6% was obtained, which means the detection of correct empty and occupied park spaces. The incorrect detections were mainly due to strong shadows, cars that had dark colors and are under a strong shadow, trucks occluding the cars and cars incorrectly parked. As the goal is to run the algorithm in a low cost platform, it took 0.7s per frame in the Raspberry Pi 2, 1.7s in the Raspberry Pi model B, and less than 0.05s in an i7 PC.

4 Conclusion and Future Work

The system presented is quite simple, but that was the initial proposal, it has a success rate of 99.6%. The proposed algorithm is suitable to deal with partial car occlusions, and it works efficiently with the Raspberry Pi 2 and a HD webcam. The cases of failure have a reduced impact and the system recovered in a few seconds in the tests done. Regarding future developments, the system has to be tested with winter conditions, more specifically snow, with total occlusions and improved in the presence of strong shadows. The development of a web platform and an entry panel for the park are future developments for the system.

Acknowledgments

This work was supported by the FCT projects LARSyS: UID/EEA/50009/2013. The 1st author also extends thanks to Prof. Vladimir Pavlicek, from the University of West Bohemia for presenting him to this topic.

References

[1] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[2] S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 1937–1944. IEEE, 2011.

[3] Smart Parking. Smart parking - a global parking business. Online, 2015. URL `http://www.smartparking.com/`.

[4] V. Reddy, C. Sanderson, and B. Lovell. Improved foreground detection via block-based classifier cascade with probabilistic decision integration. *IEEE Tr. Circuits and Systems for Video Technology*, 23 (1):83–93, 2013.

[5] SFPark. Sfpark - about. Online, 2015. URL `http://sfpark.org/about-the-project/`.

[6] K. Yamada and M. Mizuno. A vehicle parking detection method using image segmentation. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 84(10):25–34, 2001.

[7] R. Yusnita, F. Norbaya, and N. Basharuddin. Intelligent parking space detection system based on image processing. *Int. J. of Innovation, Management and Technology*, 3(3):232–235, 2012.

Cattle identification based in biometric features of the muzzle

Marta Monteiro
marta.r.f.monteiro@hotmail.com
Vasco Cadavez
vcadavez@ipb.pt
Fernando C. Monteiro
monteiro@ipb.pt
Polytechnic Institute of Bragança, ESTiG,
Bragança, Portugal

Abstract
Cattle identification has been a serious problem for breeding association. Muzzle pattern or nose print have the same characteristics with the human fingerprint which is the most popular biometric marker. The identification accuracy and the processing time are two key challenges of any cattle identification methodology. This paper presents a robust and fast cattle identification scheme from muzzle images using Speed-up Robust Features matching. The matching refinement technique based on the matching orientation information has been proposed to eliminate the miss-matched outliers.

1 Introduction

The importance of animal identification has been considered since a long time ago in applications such as cattle classification, cattle tracking from birth to the end of food chain, and understanding animal diseases trajectory and population. Now a robust cattle identification method is an important part for consumers and food industry since the usage of robust cattle identification is related to traceability and registration for breeding and marketing. Such systems contribute not only to food safety but also to quality assurance. They help to control the spread of animal disease, reduce losses of livestock producers due to disease presence, minimize expected trade loss, and decrease the government cost of control, intervention and eradication of the outbreak diseases.

Individual animal identification could be achieved either by mechanical, electronic, or biometric methods [4]. The mechanical methods (ear tags, branding or tattoos) are invasive methods and they are not good enough for traceability purposes. Electronic-based methods mainly use external tags, RFID tags, to recognize animal. However, the use of these tags is not able to provide enough reliability to the cattle identification due to theft, fraudulent and duplication. Biometric-based methods (iris scanning and DNA analysis) are also used for animal identification. Generally speaking, the biometric methods could give high identification rates, but they are intrusive for the animals and not cost-effective compared to image processing approaches.

The muzzle pattern that is correlated with human fingerprints has been considered as a biometric marker for cattle and could be used in identification of bovine animals. Recent reports on this aspect indicate that it would play a vital role and may serve as an efficient tool for identification and correct breed differentiation.

The muzzle pattern can be captured into digital format in two ways. The first one is lifted on paper data [3] and the second one is the muzzle photo [1]. Minagawa et al. [3] used the joint pixels on the skin ridges as a key feature for muzzle print matching. Some long preprocessing steps were conducted to extract the joint pixels. Two joint pixels are matched if they are in a range of 11×11 pixel region centred in the joint pixel which is considered as the ground truth. Barry et al. [1] used eigenface algorithm for cattle identification which is originally used for human face recognition. Noviyanto and Arymurthy [5] applied Scale Invariant Feature Transform (SIFT) on muzzle print images for enhancing the identification accuracy. They use SIFT to detect the key points in the image which are the distinctive points of a muzzle pattern image. The key points are then used to match each other and the number of matched key points will be used as a measure of the pattern similarity. Tharwat et al. [6] proposed a cattle identification approach that makes use of Local Binary Pattern (LBP) to extract local invariant features from muzzle print images. They also applied different classifiers including Nearest Neighbour, Naive Bayes, SVM and KNN for cattle identification.

In this research, the muzzle photos were used as the input data for automatic cattle identification. In analogy to the human fingerprints, cattle muzzle images have some discriminative features according to the ridges and beads structures and their arrangements play an important role in designing patterns on the muzzle.

Applying the image analysis techniques of filtering and segmentation, the pixels equivalent to the centroids of the beads were extracted as key features to SURF approach which is an object recognition based method that has been evaluated for the automatic cattle identification purpose.

The number of matched features has been defined as the matching score. A matching refinement technique based on the key features' orientation information has been proposed to eliminate the outliers matched key features in order to increase the identification performance.

2 Materials and Methods

This section explains about the experimental scenario, the pattern recognition technique and the proposed method for cattle identification based on digital muzzle photo data.

2.1 Data Acquisition

The muzzle photos have been taken from fourteen animals. The muzzle photo of each individual has been taken five times. Basically the four muzzle photos of each individual are used for training database and the other one muzzle photo is used for the testing phase. The nose was cleaned to eliminate snot using tissues. The muzzle photos have been taken in different illumination and with different points of view.

In every muzzle photo, a rectangle region centered on the minimum line between the nostrils is taken as the region of interest (ROI). The illustration of the ROI is shown in Figure 1.

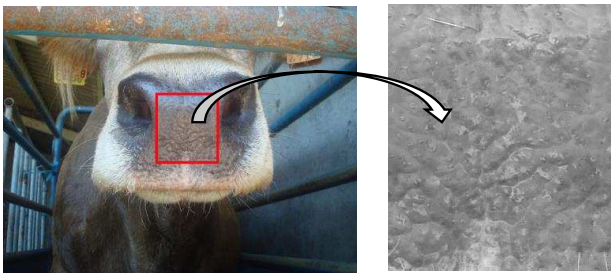


Figure 1 – The red rectangle region is the ROI of the muzzle photo. Left: original image. Right: Obtained ROI image.

2.2 Proposed Recognition Method

Speed-Up Robust Features (SURF) [2] has been claimed as a method for the object recognition which is better than its competitor, the Scale Invariant Feature Transform [5]. It is a local feature detector and descriptor that can be used for some tasks such as object recognition or registration or classification or 3D reconstruction or matching.

The proposed recognition method can be summarized as follows: (1) Collect the data set of muzzle photo; (2) Extract the ROI; (3) Extract the interest points and the descriptors using SURF algorithm for each ROI image; (4) Find the best corresponding interest points for every pair of testing image and database image; (5) Remove outliers based on the correspondence orientation between features; (6) Calculate the number of correspondences for every pair of testing image and database image; (7) Define the identification result by finding the maximum correspondence score.

2.3 Experimental Results and Discussion

In this section we will present the results obtained by the proposed method using the test image shown in Figure 1 and two training images from the database, presented in Figure 2.

The matching process is initiated by detecting the SURF features in each image. Figure 3 shows the SURF features detected in testing image and the training image from the same animal.

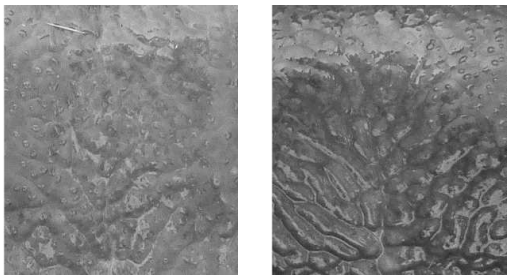


Figure 2 – Two training images from the database. Left: Image from the same animal as the testing image. Right: Image from a different animal.

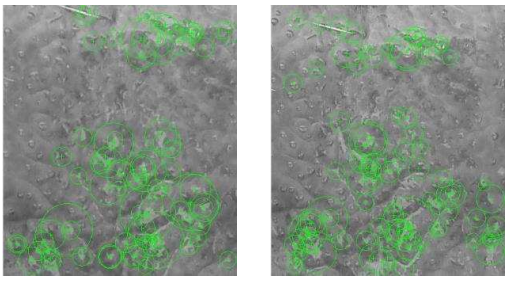


Figure 3 – SURF features detected in testing and training image from the same animal.

The next step is to extract the feature descriptors at interest points in both muzzle images. Descriptors are derived from pixels surrounding each feature point. They are needed to describe and match features specified in a single feature location.

In the matching process we find putative point correspondences using feature's descriptors. Figure 4 shows the matched features.

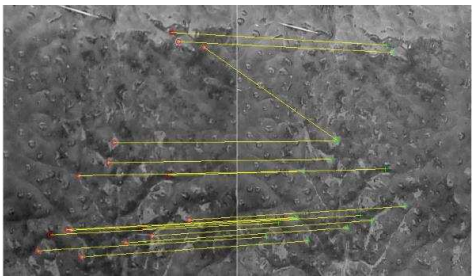


Figure 4 – Putatively matched SURF features. The yellow lines indicate the matched features

The experimental results show that the SURF still produces outliers which lead to increase the number of matched features in the matching process. In order to eliminate the matching outliers we propose a refinement process based in the features matching orientation.

First, we obtain the slope of each line correspondence. Then it was calculated the mode and the standard deviation of the slopes. Finally, it was considered that all matched features which have a slope that were out of the range $\text{mode} - std \leq slope \leq \text{mode} + std$ would be eliminated. Figure 5 shows the matching refinement result.

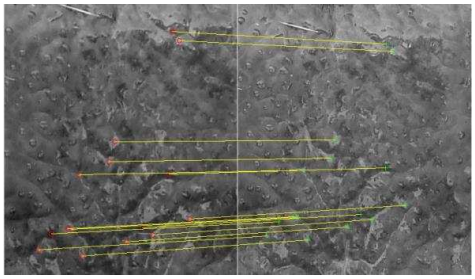


Figure 5 – Matched refinement result.

The number of correspondences between the two images from the same animal was 26.

Figure 6 shows the SURF features detected in the testing image and the training image of a different animal.

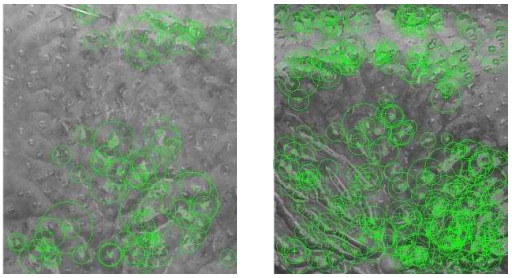


Figure 6 – SURF features detected in testing and training image of a different animal.

In the matching process we tried to find point correspondences using feature's descriptors, however the process did not find any matching feature between the two images as shown in Figure 7. With a score of 0 the automatic identification system indicates that it is a different animal.

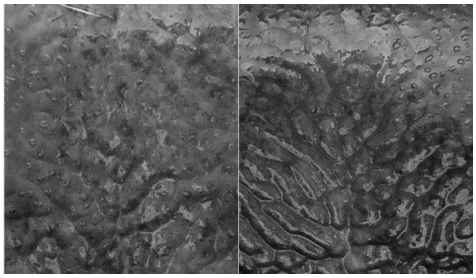


Figure 7 – Matched SURF features between testing and training images of a different animal

3 Conclusion

The points extracted from the muzzle pattern images are a good feature for the cattle identification problem especially to handle noisy data. The SURF approach and the proposed matching refinement technique can be a potential method for the beef cattle identification based on the photo image of the muzzle pattern. We obtained a 100% correct identification in all the fourteen experiments.

References

[1] B. Barry, U. A. Gonzales-Barron, K. McDonnell, F. Butler, S. Ward. Using Muzzle Pattern Recognition as a Biometric Approach for Cattle Identification. *American Society of Agricultural and Biological Engineers*, 50(3):1073-1080, 2007.

[2] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool. Speeded-Up Robust Features (SURF), *Computer Vision and Image Understanding*, 110(3):346-359, 2008.

[3] H. Minagawa, T. Fujimura, M. Ichiyanagi and K. Tanaka. Identification of beef cattle by analyzing images of their muzzle patterns lifted on paper, *Publications of the Japanese Society of Agricultural Informatics*. Vol.8, pages 596–600, 2002.

[4] C. Shanahan, B. Kernan, G. Ayalew, K. McDonnell, F. Butler and S. Ward. A framework for beef traceability from farm to slaughter using global standards: An irish perspective. *Computers and electronics in agriculture*, 66(1):62-69, 2009.

[5] A. Noviyanto and A. M. Arymurthy. Beef cattle identification based on muzzle pattern using a matching refinement technique in the SIFT method. *Computers and Electronics in Agriculture*, Vol. 99, pages 77-84, 2013.

[6] A. Tharwat, T. Gaber, A. E. Hassanien, H. A. Hassanien and M. F. Tolba. Cattle identification using muzzle print images based on texture features approach. In *Proc. Int. Conference on Innovations in Bio-Inspired Computing and Applications*, pages 217-227, 2014.

An improved method for juxta-vascular nodule candidate detection

Tiago Freitas
bio11051@fe.up.pt
INESC TEC and Faculty of Engineering, University of Porto
Porto Portugal

Aurélio Campilho
campilho@fe.up.pt
INESC TEC and Faculty of Engineering, University of Porto
Porto, Portugal

Abstract

In this paper we propose a new 3D Hessian based medialness filter for the candidate detection phase in order to improve the quality of the juxta-vascular nodules that were identified as a problem in some recent approaches.

The performance of the proposed method was tested on 115 scans from LIDC-IDRI public database, containing 360 nodules. Our approach shows a significant improvement for the juxta-vascular cases, by having a considerable reduction on the number of false positives (FP) when comparing with other methods.

1 Introduction

American Cancer Society estimated that in the year of 2014, 27% of all cancer deaths correspond to lung cancer cases. Computer Tomography (CT) is the medical imaging modality mostly used for diagnosis where it is crucial to have an early and correct detection of malignant nodules.

Computer-Aided Systems (CAD) systems can play an important role in helping the radiologists to detect the more suspicious cases. Typically, CAD systems have two main sub-systems: the first one detects the nodule candidates, and the second one that tries to reduce the number of false positives (FP). The research group develops research on lung nodule detection and recently Novo *et al*. [1] proposed an approach for lung segmentation, and a nodule detection method in [2]. A master thesis on lung nodule classification was presented in [3]. This paper focus on the first phase of the CAD system, by proposing a new method that extracts the nodule candidates with the use of the eigenvalues from the 3D Hessian matrix of the CT scans. Many methods fail to detect the nodules closer to vascular structures (juxta-vascular nodules). In our approach we propose a new way for addressing this problem, achieving a better performance on these difficult cases.

2 Methodology

Our proposal is based on the estimation of the eigenvalues from the 3D Hessian matrix, $\boldsymbol{H}$, for each voxel. To address the different size of the structures, we use a range of scales and combine the results, using the maximum of the responses over the scales σ. $\boldsymbol{H}$ is calculated using a Gaussian smoothing filter, G, with scale σ, and the corresponding second derivatives. For example the second order derivative of an image I at the scale σ w.r.t. x, is given by:

$$\frac{\partial^2 I_\sigma}{\partial x^2} = I(x) * \frac{\partial^2 G(x,\sigma)}{\partial x^2} \quad (1)$$

where $G(x, \sigma)$ is the Gaussian Kernel with the scale σ. Using equation (1), we can construct the 3D Hessian Matrix in the voxel x for each scale σ:

$$H = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yz} & I_{zz} \end{bmatrix} \quad (2)$$

The double index in the subscript indicates the second order derivatives w.r.t. the indexes. We normalize the derivatives multiplying them by σ^2. From this matrix we can obtain the eigenvalues for each scale, $|\lambda_1| < |\lambda_2| < |\lambda_3|$. Using these eigenvalues we calculate the enhancement response $V_\sigma(\mathbf{x})$ for each scale an each voxel. The final response corresponds to the maximum response over a range of scales an each voxel:

$$V(\mathbf{x}) = \max_{\sigma_1 < \sigma_j < \sigma_n} V_{\sigma_j}(\mathbf{x}) \quad (3)$$

The enhancement response, $V_\sigma(\mathbf{x})$, may have different formulations. We will compare our method with the one used by Novo *et al.*[2], that uses a central adaptive medialness principle given by:

$$V(\mathbf{x},\sigma) = \begin{cases} 0, & if\ \lambda_1 + \lambda_2 + \lambda_3 \geq 0 \\ -\frac{\lambda_2}{\lambda_3}(\lambda_2 + \lambda_3), & otherwise \end{cases} \quad (4)$$

This method does not use any kind of pre-processing on the images (only uses the Gaussian smoothing in the calculus of the image derivatives) and is sensitive to small and bright objects. The authors refers that their approach shows misdetections on the juxta-vascular nodules due to the presence of tubular structures around the nodule candidate that are also detected by the filter. In the paper we use a different formulation to overcome this problem.

For nodular structures brighter than the background the three eigenvalues are negative and having similar eigenvalues due to the spherical shape of a nodule. For tubular structures, as the blood vessels, one of the eigenvalues is larger than the others, and the other two are small. Taking this into consideration, and being λ_3 the eigenvalue with the largest absolute value, our method only computes $V_\sigma(\mathbf{x})$ for $|\lambda_3| > \alpha|\lambda_1|$ and $|\lambda_3| > \alpha|\lambda_2|$, where λ_1 and λ_2 are the eigenvalues with the smallest absolute values and α is a constant, with $\alpha > 1$ (in our experiments $\alpha = 5$). This allows the removal of most of the tubular structures and have a better performance in the juxta-vascular nodules candidates.

In order to reduce the FP, The method in [2] and the method herein described divide the candidates in three groups: smaller candidates (the ones that are grown for having a better characterization), intermediate size candidates (directly taken to the final candidates) and the large candidates (this ones are applied a progressive erosion in order to try to remove tubular and other surrounding structures that we want to remove).

3 Results

The two methods were tested in 115 different CT scans with a different number of slices. These scans were obtained from the LIDC-IDRI public database [4]. These scans had a total number of 360 nodules, witch diverse specialists indicated the location of the nodules and some of its features (size, calcification, malignancy, etc.). The candidates were considered as True Positives if they had been detected in a distance less or equal to 3 voxels from the centre of the ground truth indicated by the experts. We compared our results with the obtained by Murphy [5] and Novo *et al* [2] in the same images. It is important to refer that the results obtained by these authors were obtained removing the non-solid nodules as referred in [2].

The results of the approaches are shown in figure 1, where we test the algorithms in a synthetic juxta-vascular model, and in figure 2 and 3 where we observe the performance of both algorithms in the presence of a juxta-vascular nodule in real CT scans. The objects retained are clearly improved in our algorithm.

As we can see our method is able to remove most of the tubular structures surrounding the nodule with a considerable reduction on the number of false positives. The sensitivity and final FP/scan results are shown in table 1 where the three methods are compared.

Although we obtained a better performance in most nodules and juxta-vascular nodules, our sensitivity is lower than the other two algorithms. This happens because the juxta-pleura nodules have a different behaviour and with this new proposal some of these nodules are not detected (due to a different eigenvalues distribution near the pleura). It is important to refer that contrarily to the results obtained by the other two methods, we did not remove non-solid and the nodules and the ones that created some discord between different doctors. The fail of our method is on the juxta-pleura nodules (as observed in figure 4) that must be detected by a different method.

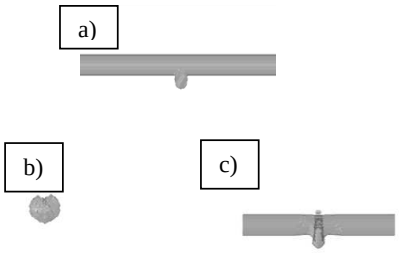


Figure 1: Isosurface of the results in synthetic juxta-vascular model obtained by our algorithm (b) and by Novo's algorithm (c).

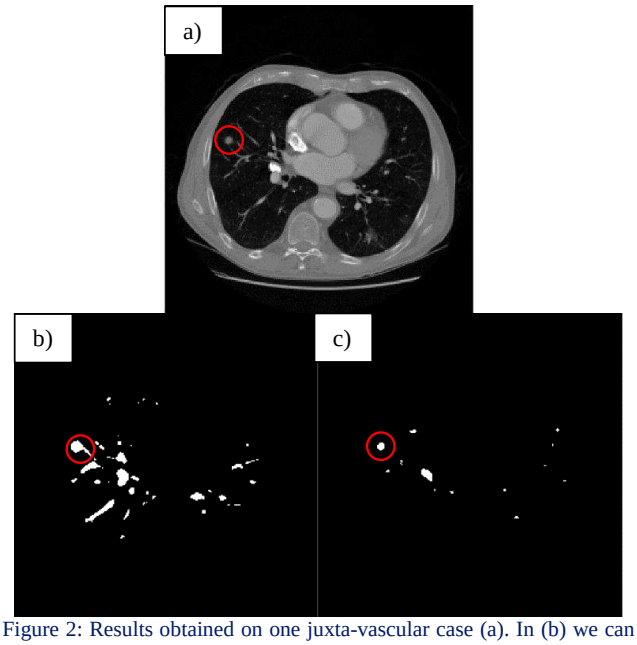


Figure 2: Results obtained on one juxta-vascular case (a). In (b) we can see the candidates (the white blobs represent the nodule candidates) obtained by Novo *et al.* method and in (a) by our new proposal.

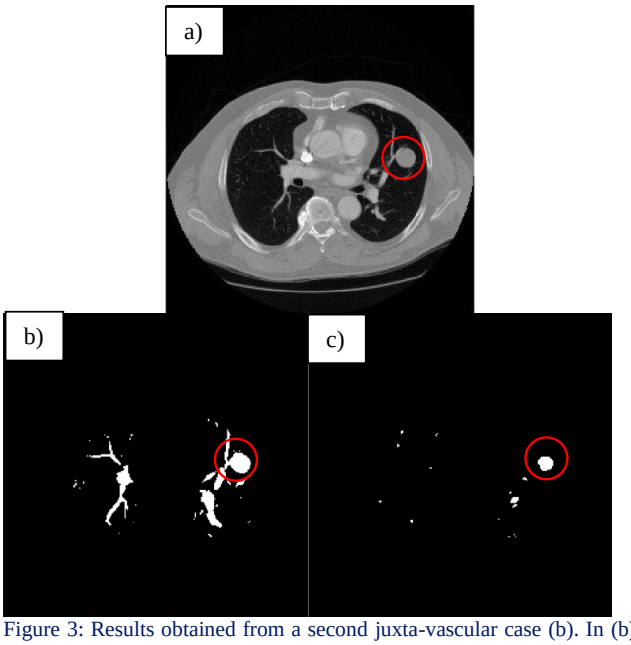


Figure 3: Results obtained from a second juxta-vascular case (b). In (b) we can see the candidates (the white blobs represent the nodule candidates) obtained by Novo *et al.* method and in (a) by our new proposal.

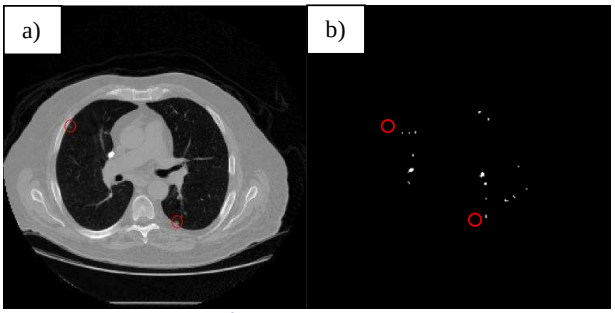


Figure 4: Results obtained for two juxta-pleura cases. In (a) we have the original image with the location of the nodules and in (b) we see that those are not detected by our method.

Table 1: Results after nodule candidates detection (Novo and Murphy Methods removed non-solid nodules and we did not did that distinction).

Methodology	Sensitivity	FP/Scan
Novo *et al.* [2]	96.22%	881
Murphy *et al.* [5]	98.11%	397
Proposed	83.34%	380

Table 2 illustrates the results on a dataset without the juxta-pleura nodules, where some misdetections occur. This clearly show an improvement of sensitivity, which is a good indication that a special treatment on the justapleural nodules is needed.

Table 2: Results obtained by our method in a dataset without juxta-pleura nodules.

Method	Sensitivity	FP/Scan
Proposed	92,62%	380

4 Conclusion

In this paper we propose a new method for nodule candidate detection based on a 3D Hessian eigenvalues medialness filter in order to improve the results on juxta-vascular nodules. We obtained a significant improvement in most nodules but with a loss in the sensitivity to juxta-pleura nodules. The false positives per scan were also improved by our method when compared to Novo *et al* [2] and Murphy *et al* [5] algorithms.

Most of the non-detected nodules are from non-solid nodules that have a different behaviour and from nodules that are very difficult to observe in the images.

Future work will include the combination of our method with an algorithm for juxta-pleura nodules detection and the development of a method for feature extraction and classification to improve the results and to reduce the number of final FP/scan.

References

[1] Jorge Novo, José Rouco, A. Mendonça, Aurélio J. C. Campilho: Reliable Lung Segmentation Methodology by Including Juxtapleural Nodules. ICIAR (2) 2014: 227-235

[2] Jorge Novo, Luis Gonçalves, Ana Maria Mendonça, Aurélio Campilho, 3D Lung Nodule Candidates Detection in Multiple Scales, IAPR MVA 2015, International Conference on Machine Vision Applications, pp. 61-64, 2015.

[3] Luís Gonçalves, Automatic Lung Nodule Classification in Chest Computerized Tomography Images, Master Thesis on Biomedical Engineering, Faculdade de Engenharia, Universidade do Porto, 2015

[4] S.G. Armato, G. McLennan et al.: "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans", Medical Physics, 38, 915–931, 2011.

[5] K. Murphy, B. van Ginneken et al.: "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and k-nearest-neighbor classification", Medical Image Analysis, 13,757–770, 2009.

Normalized relative compression (NRC) in ECG biometric identification

Susana Brás
susana.bras@ua.pt
Armando J. Pinho
ap@ua.pt

Instituto de Engenharia Electrónica e Informática de Aveiro (IEETA), DETI
Universidade de Aveiro
Aveiro, Portugal

Abstract

In this work, it is presented a new measure for ECG-based biometric systems. To accomplish this goal, we use a similarity measure that relies on the concept of Kolmogorov complexity. The proposed similarity measure allows the comparison of two ECG segments, without having to follow traditional approaches that require heartbeat segmentation (described as highly influenced by external or internal interferences). The proposed measure relies on the concept of conditional compression of two objects, *i.e.*, the compression of one object performed considering *exclusively* the data of the other object. And therefore, it is expected that when we compare two ECG segments from the same participant, it is expected to obtain smaller values than when we compare ECG segments from different participants. We observed that the proposed measure is able to correctly differentiate between the three studied groups: 1) comparing the same ECG segment: 0.150 ± 0.103; 2) comparing two different ECG segments from the same participant: 0.347 ± 0.097; 3) comparing two ECG segments from different participants: 0.568 ± 0.135.

1 Introduction

Biometric identification refers to methods that use biological data in order to identify someone as being himself. Considering these methods there are applications using iris, face, fingertip, electrocardiogram (ECG), electroencephalogram (EEG) [1, 7]. Some of these methods are susceptible to be falsified since there is no need of contact or liveliness for acquisition [1]. Our heart is continuously beating and possesses an inter-subject variability that may be used to obtain discriminant information based on the electrical heart characteristics.

The ECG is a well known signal in biomedical applications. Usually it is intended to reduce the inter variability that characterizes this kind of signals. Nevertheless, is precisely this inter variability that renders the ECG an interesting signal on biometric systems [1].

There is an increasing interest in using the ECG in biometric identification [1, 2, 7]. The ECG presents changes due to the circadian cycle, or some particular circumstances (*e.g.*, stress, fatigue), conducing to alterations in rhythm and/or amplitude [7]. To overcome this shortcoming, the ECG biometric applications should correctly deal with fluctuations and noise [1, 7].

The ECG biometric methods are usually divided in two groups: fiducial and non fiducial [7]. The fiducial methods use as pre-processing step the heartbeat delineation (identification of specific points that constitutes the ECG) [4] which is susceptible to processing errors, due to (*e.g.*) noise [1, 7]. The non fiducial methods do not need the ECG wave delineation or alignment, and therefore it is expected that these methods reduce the pre-processing error [1, 7].

In order to obtain a feasible and accurate biometric system, it is needed to implement methods that may correctly deal with traditional problems in biometric identification methods. In this work, we propose the use of parameter free data mining methods, in particular the use of compression methods that are based on Kolmogorov complexity, for ECG biometric identification.

2 Methods

For method development, it was used the control subjects from the PTB database [3]. The ECG signal was decimated to 500 Hz, using an eighth-order low pass Chebyshev Type I filter with a 200 Hz cutoff frequency.

The baseline wander was filtered using a moving average filter with window width of 100 samples. A Butterworth low pass filter of order 10 with a cut off frequency of 40 Hz was also applied.

Each participant's record in the dataset was split in five segments of 20 seconds, constituting a final database of 52 participants and 260 records.

2.1 Quantization

When compression algorithms are used, it is generally preferable the use of symbolic records. Therefore, the ECG should be converted from the real-valued numerical signal to a symbolic time series. To accomplish this step we implemented a quantization method, the SAX algorithm [6]: the time series is normalized and divided in N segments of dimension w. For each N_i segment, the method calculates its mean value, which will match a symbol in the new data dimension. The method optimization will be dependent on the alphabet size and on the w dimension of the series segments.

2.2 Compression Method

The use of Kolmogorov theory allows the definition of similarity measures between binary objects. In order to approximate the Kolmogorov complexity, a compression algorithm needs to be able to accumulate knowledge of the data while the compression is performed, *i.e.* it is needed to compute a model of the data. Finite-context models are an example of methods that represent data information [9]. In the context of similarity/dissimilarity measures, these models had been proved to be efficient, relying on the data algorithmic entropy [8].

The specific measure that we applied is based on the notion of relative compression of two objects, i.e., the compression of one object done *exclusively* using the information of the other object. We define the normalized relative compression (NRC) of x given y as

$$\mathrm{NRC}(x,y) = \frac{C(x||y)}{|x|}, \tag{1}$$

where $|x|$ is the size of the object. Basically, this measure gives information about the amount of data in x that cannot be described by y. The $C(x||y)$ uses a combination of finite-context models of several orders (k) to build an internal model of y, which is kept fixed afterwards. Then, x is encoded exclusively using the model built from y. In order to estimate such models, three parameters had to be considered: k, γ, and α (for details on the encoding method, please see [8, 9])

2.3 Statistical Analysis

In method evaluation we used a leave one out strategy. Four segments from each participant were used as training dataset, and the remaining segment was used as test dataset in the NRC internal model development.

In order to evaluate the measure performance in the biometric identification, it is intended to evaluate the NRC measure between two ECG segments. Therefore, we grouped the NRC results from the ECG comparisons aggregating the self-evaluation (NRC value of the record with itself), the participant measure (NRC value calculated between records of the same participant, excluding self-similarity), and the out measure (NRC value calculated between records of different participants). The results were presented as mean $\pm$ standard deviation of the NRC.

3 Results and Discussion

The dataset was divided in train (first 10 participants - 50 ECG segments) and test datasets, in order to optimize the method's parameters. Considering the quantization an alphabet size of 20 and a window size of 3 points (which will correspond to 1 symbol in the new data dimension) were found to be optimal. In the NRC implementation a γ of 0.2 and a combination of (α, k) pairs taken from set $\{(1/1000, 5), (1/1000, 6), (1/1000, 7), (1/1000, 8), (1/1000, 9)\}$ were used. In order to find the

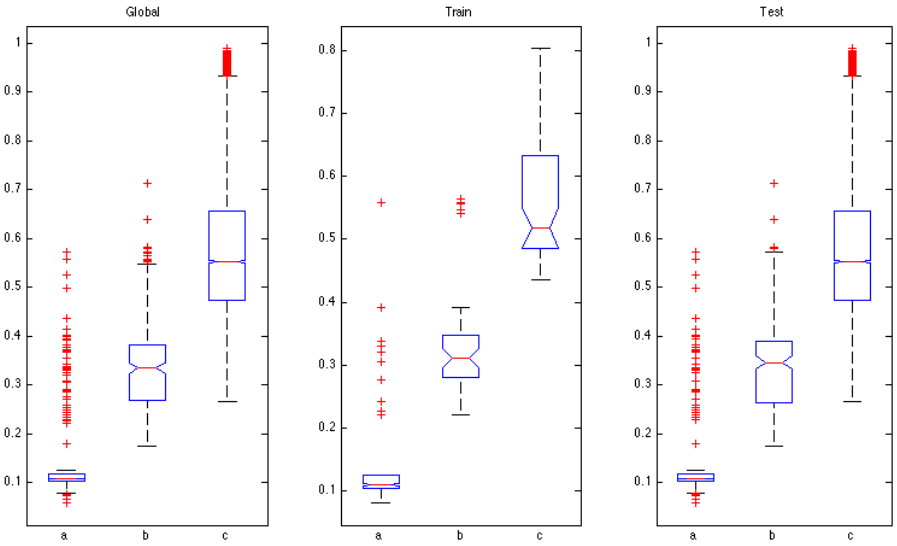


Figure 1: Boxplots representing the NRC measure evaluating data from the global, the train and the test datasets. The *a* box represents the self-evaluation, the *b* represents the participant measure, and the *c* represents the out measure.

Table 1: Summary of the results obtained using the NCR measure over the train, test and global datasets. The results are presented as $mean \pm std$ of the calculated measure.

Dataset	self-evaluation	participant measure	out measure
Train	0.149 ± 0.096	0.325 ± 0.087	0.563 ± 0.102
Test	0.151 ± 0.104	0.346 ± 0.099	0.568 ± 0.135
Global	0.150 ± 0.103	0.347 ± 0.097	0.568 ± 0.135

optimal parameters several tests were performed. It was considered the minimization of the self-distance (comparing the record with itself) as objective function.

In this work, it is intended to evaluate the ability of the measure in the identification of ECG segments from the same participant. To accomplish this goal the NRC was calculated between ECG segments from the same or different participants. In Figure 1 the boxplot evidenced the measure ability in the discrimination between the three proposed tests: self-evaluation (*a*), participant measure (*b*), out measure (*c*). Table 1 presents the NRC values, evidencing that when ECG segments from different participants are compared, the measure shows higher values than when ECG segments from the same participant are compared.

The standard deviation of NRC should be reduced in the test dataset, which may be related with database, or even model optimization process. Considering that the ECG is susceptible to alterations, the NRC measure should be validated using a database that comprises ECG data collected in different days and subject to external variability conditions.

The use of parameter free data mining methods was proved to be efficient in classification, since there is no pre-assumption about the premises, which allows true exploratory data mining [5]. The NRC measure is an example of a parameter free data mining. Moreover, since there is no need of heartbeat segmentation, the pre-processing error associated with this task is reduced. These results indicate that the NRC measure may be feasible in biometric identification tasks.

4 Acknowledgment

This work was supported by National Funds through FCT - Foundation for Science and Technology, in the scope of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014, (IEETA/UA), and CMUP-ERI/FIA/0031/2013 (VR2Market, by CMU-Portugal program). S. Brás acknowledges the Postdoc Grant from FCT, ref. SFRH/BPD/92342/2013.

References

[1] Foteini Agrafioti and Dimitrios Hatzinakos. ECG biometric analysis in cardiac irregularity conditions. *Signal, Image and Video Processing*, 3(4):329–343, 2009. ISSN 1863-1703.

[2] David Pereira Coutinho, Hugo Silva, Hugo Gamboa, Ana Fred, and Mário Figueiredo. Novel fiducial and non-fiducial approaches to electrocardiogram-based biometric systems. *IET biometrics*, 2(2): 64–75, 2013. ISSN 2047-4946.

[3] Ary L Goldberger, Luis A N Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch. Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, 2000. ISSN 1524-4539.

[4] Steven A Israel, John M Irvine, Andrew Cheng, Mark D Wiederhold, and Brenda K Wiederhold. ECG to identify individuals. *Pattern recognition*, 38(1):133–142, 2005. ISSN 0031-3203.

[5] Eamonn Keogh, Stefano Lonardi, and Chotirat Ann Ratanamahatana. Towards parameter-free data mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 206–215. ACM, 2004. ISBN 1581138881.

[6] Jessica Lin, Eamonn Keogh, Stefano Lonardi, and Bill Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pages 2–11. ACM, 2003.

[7] Ikenna Odinaka, Po-Hsiang Lai, Alan D Kaplan, Joseph A O'Sullivan, Erik J Sirevaag, and John W Rohrbaugh. ECG biometric recognition: A comparative analysis. *Information Forensics and Security, IEEE Transactions on*, 7(6):1812–1824, 2012. ISSN 1556-6013.

[8] Armando J Pinho and Paulo Jorge S G Ferreira. Image similarity using the normalized compression distance based on finite context models. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1993–1996. IEEE, 2011. ISBN 1457713047.

[9] Armando J Pinho, Paulo J S G Ferreira, António J R Neves, and Carlos A C Bastos. On the representability of complete genomes by multiple competing finite-context (Markov) models. *PloS one*, 6(6): e21588, 2011. ISSN 1932-6203.

A 3D Parametric Model for Breast Data

Hooshiar Zolfagharnasab
dee12012@fe.up.pt
Jaime S. Cardoso
jaime.cardoso@inescporto.pt
Hélder P. Oliveira
helder.f.oliveira@inescporto.pt

Faculdade de Engenharia, Universidade do Porto
Rua Doutor Roberto Frias, S/N, 4200-465
Porto, Portugal
INESC TEC
Campus da FEUP, Rua Doutor Roberto Frias, 4200-465
Porto, Portugal

Abstract

Breast cancer is the most common cancer disease among females. Therefore any deformations in the breast shape resulted from treatment, impacts patients' quality of life given the importance of breast as a feminine symbol. Using techniques such as parametric modeling, enable surgeons to model 3D breast shape virtually. Such parametric model can be used in Surgery Planning Tools in order to follow up the shape of breast after applying different deformation. This framework can help not only to improve the surgical skills of surgeons to perform surgeries with better cosmetic outcomes, but also to increase interaction between patients and surgeons in the moment of discussing what procedure needs to be performed. In this paper, two different methodologies of parametric modeling are compared. Quantitative analysis indicates that Free-Form-Deformation methodology (FFD) presents better parametric models than Physical Modeling methodology.

1 Introduction

Nowadays, breast cancer is the most common cancer among women accounting for near %23 of all cancers. The tumor characteristics (including position and size of tumor, diagnosing time, etc.) defines the appropriate surgery approach [8]: mastectomy in which total tissues of the breast are removed, resulting in huge deformation of the breast, and lumpectomy which is known as BCCT in which just the tumor and a thin layer of healthy surrounding tissue are removed. Similar survival rates of both surgeries motivate surgeons to perform BCCT to prevent major breast deformations [5].

Recent researches have been focused on the effectiveness of Planning Tool on the cosmetic outcome of surgeries [9]. The initial step of a planning tool called parametric modeling, transforms the input data into a mathematical model. In this work, we will test two different methodologies of parametric modeling proposed in [6] and [2].

The main objective of a parametric model algorithm is the conversion of a set of 3D points into a mathematical model which expresses the breast shape. Thus, surgeons can modify this parametric model with different deformations in order to define new breast shapes.

The rest of this paper is structured in 4 more sections. Within section 2, required background of parametric modeling will be studied. While the availability of FFD in parametric modeling is studied in section 3, its implementation and comparison with a state-of-the-art algorithm is discussed in section 4. Finally, overall discussions are wrapped up in section 5, as conclusion.

2 Background

The main requirement of a planning tool can be fulfilled with a parametric model. Highlighting such necessity, Balaniuk *et al*. [1] combined both virtual reality approaches and soft tissue modeling methods to simulate reconstruction or augmentation surgeries using 3D tools.

Later, Ruiz *et al*. proposed a method to create a surface to be fitted to data by satisfying the Nyquist-Shannon criteria. They came to solve the minimization problem (which is defined based on residual error between the surface and the data,) is solved by the Gauss-Newton iterative method to approach the surface to the data iteratively [7].

In many computer vision problems, it is tried to find a smooth surface which can be fitted to an unstructured data. the use of NURBS is one of that possible solutions. Besides, it can fulfill the requirement of parametric modeling [4]. Beginning with the Physical Deformation methodology by [6], we briefly explain the application of fitting algorithms in parametric modeling. Then in the following, the idea of using FFD which is proposed in [2], will be studied.

2.1 Pyisical Deformation

The fitting methodology is another approach within the purpose of parametric modeling. In this regard, Physical Deformation methodology (PD) [6] is proposed to perform the fitting by changing implicit parameters of a curve (or a surface) to be fitted to the data. Within this algorithm, a superquadric model is adopted, with parameters to model five major features of the shape of the breast, including sagging, horizontal deviations, and convexity/concavity of the top half of the breast. Initial implicit model parameters are set up with experimental values to describe deformation of a typical breast. During an iterative fitting procedure, the initial values are changed in order to decrease the distance between model and breast data.

2.2 Free-Form Deformation

Starting from a superquadrics model, the algorithm in [2] first tries to fit a superellipsoid model to an input 3D breast data, and then by using FFD, it deforms the fitted superellipsoid to be as similar as it can be (based on optimization criteria) to the input data.

Defined in the implicit format, superquadrics form a family of implicit quadric surfaces.

$$F(x,y,z) = (((\frac{x}{a_1})^{\frac{a}{\varepsilon_2}} + (\frac{y}{a_2})^{\frac{a}{\varepsilon_2}})^{\frac{\varepsilon_2}{\varepsilon_1}} + (\frac{z}{a_3})^{\frac{2}{\varepsilon_1}})^{\frac{\varepsilon_1}{2}} \tag{1}$$

Considering the parameterization of a sphere in spherical coordinate, the function can be written as:

$$\begin{cases} x = \cos\eta\cos\omega & -\frac{\pi}{2} < \eta < \frac{\pi}{2} \\ y = \sin\eta\cos\omega & -\pi < \omega < \pi \\ z = \sin\eta \end{cases} \tag{2}$$

If the steps between the grids of η and ω are constant, points are distributed uniformly on the sphere. Then, in two steps, it is tried to redefine the function of superellipsoid from the sphere, so that the distribution of points are kept uniform after redefinition. Considering these steps, conversions from an square to ellipsoid and then superellipsoid can be formulated as:

$$x_e = a_1 x; y_e = a_2 y; z_e = a_3 z \tag{3a}$$
$$x_s = \rho x_e; y_s = \rho y_s; z_s = \rho z_s \tag{3b}$$

new formulation can be defined as:

$$\rho = ((((|\cos\eta\cos\omega|)^{\frac{a}{\varepsilon_2}})^{\frac{\varepsilon_2}{\varepsilon_1}} + (|\sin\eta\cos\omega|)^{\frac{a}{\varepsilon_2}})^{\frac{\varepsilon_2}{\varepsilon_1}} + (\sin\eta)^{\frac{2}{\varepsilon_1}})^{\frac{\varepsilon_1}{2}} \tag{4}$$

Calculating ρ, x_s , y_s and z_s regarding to the values of the determined domain over $[-\frac{\pi}{2},\frac{\pi}{2}]\times[-\pi,\pi]$ will result in new parameterization of the superellipsoid.

Benefiting from the light computation, the fitting algorithm of [2] aims to change the superellipsoid parameters in order to minimize the Euclidean distance between the input data and the model. Since there might not be a specific set of parameters to fit the quadric model on the point cloud completely, such problem can be categorized as a least-square issue. Defining $\hat{F}$ as the function to generate the superellipsoid, the claim of $\hat{F} = 1$ can be considered, thus the best set of parameters to minimize the equation in 5 can be computed using the derivative of E.

$$E(A) = \sum_{i=1}^{N} [1 - \hat{F}(x_d, y_d, z_d, a_1, a_2, a_3, \varepsilon_1, \varepsilon_2, \phi, \theta, \psi, t)]^2 \tag{5}$$

This equation can be presented in an easy form such as:

$$X = BP \tag{6}$$

Where P is a matrix containing the control points of the grid, X is parametric model points and B is the deformation matrix. Deformation of superellipsoid is followed by relocating the control points, shown in Figure 1, at the beginning and end of each edge of grid. Considering the δX as the displacement field between original point cloud and superellipsoid, using a linear equation system we can write:

$$\delta X = B\delta P \tag{7}$$

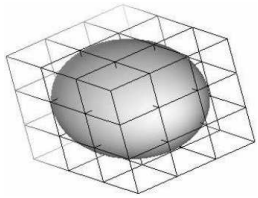


Figure 1: Superellipsoid surrounded by control grid

Within iterative steps, the points on the superellipsoid are approached to the original point cloud by solving following minimization equation:

$$min_p||BP - X||^2 := min_{\delta P}||B\delta P - \delta X||^2 + \alpha \sum_{j=1}^{NP} \sum_{j'} ||P_j - P_{j'}||^2 \tag{8}$$

Where j' corresponds to the neighbors of P and NP indicates the number of control points. The second term is an internal energy corresponding to the insertion of a zero-length springs between control points, that is being regularized by the weight of α. Such equation can be solved using singular value decomposition of the matrix B.

3 Implementation and Results

The two discussed methods in [2] and [6] were implemented using OpenCV in C++ and Mathworks MATLAB 2015a, respectively. Also the methods were compared based on Error Distance between input data and the model. It is important to notice that bi-directional error distances are reported hence the two sets contained different number of points. We used a dataset of 70 3D breast models, obtained by the 3D reconstruction algorithm proposed in [3]. The mentioned 3D reconstruction algorithm is based on a rigid ICP method to reconstruct 3D breast models from Microsoft Kinect data. Average fitting errors (Euclidean and Hausdorff) together with standard deviations are reported in Table 1. Also average number of iterations to reach to stop criterion is reported for each algorithm. Note that since the compilers of each implementation is different, time comparison cannot lead to exact conclusion.

Reported results indicate that the FFD approach performs better fitting than methodology of [6] since it presents smaller error distances, while number of required iterations are almost the same. The reason should be searched in the number of parameters. While in the methodology of [6], at most 20 parameters have been defined to be changed iteratively, methodology of FFD performs the fitting with number of parameters equal to the number of control points. In the current implementation, a cube of $7 \times 7 \times 7$ points were defined, which provides enough flexibility for FFD method to fit initial superellipsoid to the data, comparing to at most 20 parameters of Physical Deformation methodology.

Besides numerical analysis, visual comparisons are available in Figure 2, indicating better fitting based on methodology of [2].

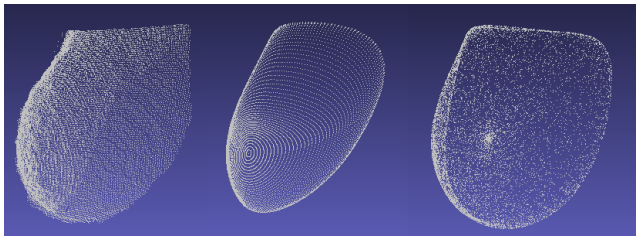


Figure 2: Visual comparison of two different parametric model reconstructions; Left: Initial breast pointcloud, middle: parametric model of methodology proposed in [6], right: parametric model of the proposed methodology in [2]

4 Conclusion

Putting all together, parametric modeling is a technique which converts input data into a mathematical model. Mentioning the importance of a parametric modeling in a planning tool, two methodologies (FFD and Physical Model) have been studied and it was shown FFD presents better results based on both numerical and visual analyses.

Acknowledgment

This work is financed by the European Community's Seventh Framework Programme [grant number FP7600948] and by Fundação para a Ciência e a Tecnologia within PhD grant number SFRH/BD/97698/2013, and by the FCT within project UID/EEA/50014/2013.

References

[1] R. Balaniuk, I. Costa, and J. Mello. Cosmetic breast surgery simulation. In *Proc. the VIII Symposium on Virtual Reality (SVR)*, 2006.

[2] E. Bardinet, L. D. Cohen, N. Ayache, S. Smith, J Paul Siebert, S. Oehler, X. Ju, and A. K. Ray. A parametric deformable model to fit unstructured 3d data. *Computer Vision and Image Understanding*, 71(1):39–54, 1998.

[3] P. Costa, J. P. Monteiro, H. Zolfagharnasab, and H. P. Oliveira. Tessellation-based coarse registration method for 3d reconstruction of the female torso. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 301–306. IEEE, 2014.

[4] N. Leal, E. Leal, and J. W. Branch. Simple method for constructing nurbs surfaces from unorganized points. In *Proc. the 19th International Meshing Roundtable*, pages 161–175, 2010.

[5] H. P. Oliveira, J. S. Cardoso, A. Magalhães, and M. J. Cardoso. Methods for the aesthetic evaluation of breast cancer conservation treatment: A technological review. *Current Medical Imaging Reviews*, 9(1):32–46, 2013.

[6] D. Pernes, J. S. Cardoso, and H. P. Oliveira. Fitting of superquadrics for breast modelling by geometric distance minimization. In *Proc. the 8th IEEE International Conference on Bioinformatics and Biomedicine*, 2014.

[7] O. Ruiz, S. Arroyave, and D. Acosta. Fitting of analytic surfaces to noisy point clouds. *American Journal of Computational Mathematics*, 3(1):18–26, 2013.

[8] American Cancer Society. Breast cancer facts and figures 2011-2012. *American Cancer Society (ACS).*, 2012.

[9] H. Zolfagharnasab, J. S. Cardoso, and H. P. Oliveira. 3d breast parametric model for surgery planning-a technical review. In *Proc. the 22nd Portuguese Conference on Pattern Recognition*, pages 85–86. APRP, 2013.

Table 1: Bidirectional distance error (Euclidean distances) of the two compared methodologies: [2] and [6]. Note that M stands for model and input data is labled with GT.

Method		Distance Error ($M \rightarrow GT$)	Distance Error ($M \rightarrow GT$)	Hausdorff ($M \rightarrow GT$)	Hausdorff ($M \rightarrow GT$)	Time (s)	No. of Iterations	Stop Criterion (Euclidean Distance)
Physical Model	μ	0.02840	0.04000	0.15560	0.31580	12	8.7	0.0015
	σ	0.00450	0.00660	0.03170	0.07460	16		
Superellipsoid + FFD	μ	0.00131	0.00170	0.00763	0.01343	3001	10.42	0.0015
	σ	0.00008	0.00025	0.00152	0.00483	1445		

Detection and visualisation of regions of human DNA not present in other primates

Diogo Pratas[1]
pratas@ua.pt
Raquel M. Silva[2]
raquelsilva@ua.pt
Armando J. Pinho[1]
ap@ua.pt
Paulo J. S. G. Ferreira[1]
pjf@ua.pt

[1] Information Systems and Processing Group, IEETA/DETI,
University of Aveiro,
3810–193 Aveiro, Portugal

[2] IEETA,
University of Aveiro,
3810–193 Aveiro, Portugal

Abstract

Human specific regions are DNA segments that are unique or share high dissimilarity rates relatively to close species, namely primates. Their existence is important to localize evolutionary traits that are often related to novel functionality, besides its obvious discriminative ability.

We propose an unsupervised method, and an associated tool, to detect and visualise these regions. It is based on the detection of relative absent words (RAWs), using a probabilistic high depth model. The experimental results show several regions that are associated with documented human specific regions, namely centromeres and several genes, such as those related with olfact. However, it also shows several undocumented ones, that may express trends in human evolution.

1 Introduction

Relative absent words (RAWs) are sub-sequences that do not occur in a given sequence (the reference), but do occur in another sequence (the target). Consider a target sequence, x, and a reference sequence, y, both from a finite alphabet Θ. We say that β is a factor of x if x can be expressed as $x = u\beta v$, with uv denoting the concatenation between u and v. We denote by $\mathcal{W}_k(x)$ the set of all k-size words (or factors) of x. Also, we represent the set of all k-size words *not in* x as $\overline{\mathcal{W}_k(x)}$. For each k-size word, we denote the set of all words that exist in x but do not exist in y by

$$\mathcal{R}_k(x,\bar{y}) = \mathcal{W}_k(x) \cap \overline{\mathcal{W}_k(y)} \tag{1}$$

and the subset of words that are minimal as

$$\mathcal{M}_k(x,\bar{y}) = \{\beta \in \mathcal{R}_k(x,\bar{y}) : \mathcal{W}_{k-1}(\beta) \cap \mathcal{M}_{k-1}(x,\bar{y}) = \emptyset\}, \tag{2}$$

i.e., a minimal absent word of size k cannot contain any minimal absent word of size less than k. In particular, $l\beta r$ is a minimal absent word of x, where l and r are single letters from Θ, if $l\beta r$ is not a word of x, but both $l\beta$ and βr are ([6]).

Although minimal absent words have been studied before to describe properties of prokaryotic and eukaryotic genomes and to develop methods for phylogeny construction or PCR primer design [3, 9], their pratical usage for differential analysis is relatively new. Recently, we have proposed this approach, exploring the non-empty set $\mathcal{M}_k(x,\bar{y})$ corresponding to the smallest k, referred to as minimal relative absent words (mRAWs), in an application related to the ebola virus [7].

In this paper, we focus on finding large RAWs, with the aim of detecting human regions that are unique (with high probability), relatively to several primates. Hence, we are interested in creating a model of one or more reference sequences, to detect sub-sequences that are present in a target. To achieve this goal, we use a k-mer model with high depth (tipically $k = 30$), that is efficiently implemented using Bloom filters. Thereafter, the unique regions, that are filtered and segmented using a threshold value, are presented in a map. A tool, with the implementation of the unsupervised method, is freely available.

2 Method

If one uses a binary vector to store all the possible entries indicating if a certain k-mer exists or not in the sequence, we would use 4^k bits. For $k = 30$, we would need 131,072 TeraBytes of memory, which is impracticable on current computers. A data structure such as a hash table for implementing such a model would certainly be more reasonable, but the memory becomes dependent on the number of inserted elements. Moreover, for the volumes of data that we usually need to deal with, it still implies high memory requirements.

A third option is a probabilistic data structure, namely a Bloom filter [1], which trades space resources by precision. Notwithstanding, the usage of a very large Bloom filter (with the number of hash functions optimized), can give very high probabilities of becoming very similar to deterministic. Because, for this case, we do not need very large lengths and precise results, since we want to find regions (RAWs) and not mRAWs, this seems the most efficient choice.

For using a Bloom filter, we set a vector of dimension m and the number of hash functions h, obtaining a balance that is also related with the number of elements that are filtered, n. Asymptotically, for a given m and n, the value of the number of hash functions that minimizes the probability of false positives is given by

$$h = \frac{m}{n} \ln 2, \tag{3}$$

that can be re-written as

$$2^{-h} \approx 0.6185^{m/n}. \tag{4}$$

The more elements that are added to the set, the larger the probability of false positives. Given n and a desired false positive probability p (assuming that the optimal value of h is used), we can find the required number of bits m using

$$m = -\frac{n \ln p}{(\ln 2)^2}. \tag{5}$$

This means that, asymptotically, for a given false positive probability p, the length m of a Bloom filter is proportional to the number of elements being filtered, n. For finite values, the false positive probability for a finite Bloom filter with m bits, n elements and h hash functions is, at most (see [2] for more details),

$$\left(1 - e^{-h(n+0.5)/(m-1)}\right)^h. \tag{6}$$

This method allows whole genome analysis using T_n targets and R_n references. To solve this, we write to disk each RAW detected from R_i in relation to each T_i. Next, for each T_i, the RAWs are considered only if they exist in all R_i. A file containing the whole genome RAWs in relation to each T_i is stored (these are the unique regions).

An example of the method, from the sequences to the maps, using three reference sequences and one target is depicted in Fig. 1. For T_n targets, the process is repeated n times. Moreover, when using inverted repeats, the reverse complemented sequence is also loaded into memory (for the same reference model).

For visualising the unique regions, after low-pass filtering of a binary sequence containing the presence/absence of RAWs, a threshold is used to segment the regions and then they are presented in a visual map (see Fig. 1 for an example).

We have created a fully automatic tool (CHESTER), written in C language, with the implementation of the unsupervised method. It is available at `http://github.com/pratas/chester`, under GPL-2, and can be applied to any genomic sequence, in FASTA, FASTQ or SEQ (ACGTN) formats.

In Fig. 2, we show the results of running the tool against several synthetic sequences, used to better illustrate the method. As can be seen,

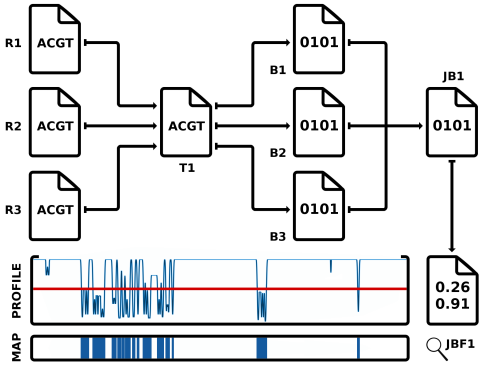


Figure 1: Visual description of the method. The genomic sequences contained in the files $R1$, $R2$ and $R3$ are independently processed against a target, $T1$. From each computation a binary sequence is generated, $B1$, $B2$ and $B3$, describing the presence/absense of a RAW according to the order of $T1$. Next, the binary files are merged using a logic or ($\vee$), $B1 \vee B2 \vee B3$ and the result is $JB1$. The $JB1$ sequence is then low-pass filtered, resulting in the real sequence described as $JBF1$. Finally, a threshold (line in red) is used to segment the information contained in the $JBF1$, where each segmented region is represented in the RAWs map.

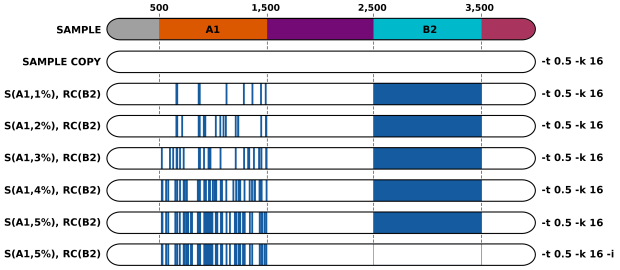


Figure 2: Running CHESTER using several synthetic sequences. Blocks $A1$ and $B2$ have been edited according to the functions refered on the left. Function S stands for a substitution mutation of the input block with the defined percentage. Function RC applies the reverse complement of the input. CHESTER running parameters are defined on the right, using a threshold of 0.5 and a k-mer size of 16, while only the bottom map has been run using inversions. The blue color on the computed maps represents the unique regions according to the RAWs.

the method identifies as novel the regions that are mutated ($A1$) and also the inversions ($B2$). When the tool runs with the "-i" parameter (handle inversions), these where successfuly not reported. When dealing with sequenced data, the sequences might have several inverted regions due to errors of assemblage or sequencing. As we have shown in Fig. 2, this method is prepared to overcome those limitations.

3 Experimental results

In the experiments with real sequences we used: reference human genome (GRC-38) [4]; reference chimpanzee genome (2.1.4); reference gorilla genome (3.1); reference orangutan genome (2.0.2). The sequences were downloaded from the NCBI. The Y chromosomes of gorilla and orangutan have not been yet sequenced and therefore they are not present. On the other hand, we have included the unlocalized, unplaced and mitochondrial sequences, in order to bypass most assembly challenges.

We ran CHESTER on those sequences, obtaining the map displayed in Fig. 3. The larger blue areas identify the centromeres, corresponding to very repetitive DNA. The smaller areas contain several genes and pseudogenes (genes that are not expressed [8]) associated, for instance, to immunology, blood, smell and brain. Besides, there are several identified motifs in these regions (on the NCBI and Ensemble) that, although considered of importance, their nature has not yet been understood.

Of those sub-sequences which are less understood, we highlight HCP5 HLA complex P5, MAFK, GALNT9, OR11H12, OR11H11, SHOX short stature homebox. For example, on human chromosome 14 the OR11H12

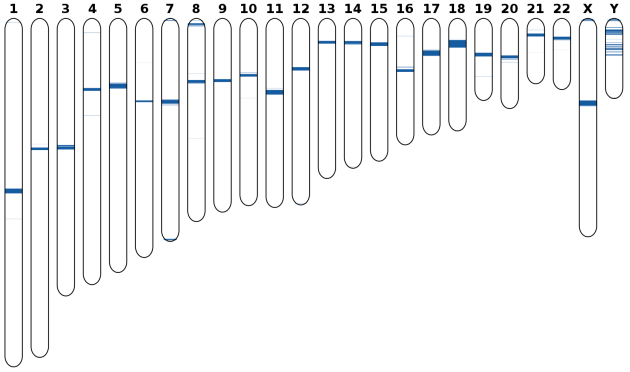


Figure 3: Human specific region (chromosomal) maps relative to the chimpanzee, gorilla and orangutan using CHESTER with $t = 0.6$ and $k = 30$. The blue strips represent the relative unique regions.

olfactory receptor is a gene associated with olfactory receptors that interact with odorant molecules in the nose, to initiate a neuronal response that triggers the perception of a smell. These findings are confirmed by other recent studies that show the loss of olfatory function only in the hominid evolution and therefore the consequent genomic sequence alteration [5].

4 Conclusions

RAWs are unique words that appear in a sequence and nowhere in other sequence. Their fundaments have been used recently in personalized medicine scenarios using minimal absent words. In this paper, we followed a different line, exploring high orders and whole genome analysis, proposing a method and an associated tool (CHESTER) to detect and visualise RAWs. These regions are associated with relative whole genome uniqueness, namely with centromeres and recent evolutionary traits that, relatively to several primates (chimpanzee, gorilla and orangutan), are specific to humans.

Acknowledgment

This work was partially funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 305444 "RD-Connect: An integrated platform connecting registries, biobanks and clinical bioinformatics for rare disease research" and by National Funds through FCT - Foundation for Science and Technology, in the context of the project UID/CEC/00127/2013.

References

[1] B. H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Commun. ACM*, 13(7):422–426, July 1970.

[2] A. Broder and M. Mitzenmacher. Network applications of bloom filters: A survey. *Internet Math.*, 1(4):485–509, 2004.

[3] S. Chairungsee and M. Crochemore. Using minimal absent words to build phylogeny. *Theo. Comp. Sci*, 450:109–116, 2012.

[4] D. Church, M. Deanna, V. Schneider, et al. Modernizing reference genome assemblies. *PLoS Biol.*, 9(7):e1001091, 2011.

[5] G. M. Hughes, E. C. Teeling, and D. G. Higgins. Loss of olfactory receptor function in hominin evolution. *PloS ONE*, 9(1):e84714, 2014.

[6] A. J. Pinho, P. J. S. G. Ferreira, S. P. Garcia, and J. M. O. S. Rodrigues. On finding minimal absent words. *BMC Bioinformatics*, 10 (137), May 2009.

[7] R. M. Silva, D. Pratas, L. Castro, A. J. Pinho, and P. J. S. G. Ferreira. Three minimal sequences found in ebola virus genomes and absent from human DNA. *Bioinformatics*, page btv189, 2015.

[8] E. F. Vanin. Processed pseudogenes: characteristics and evolution. *Annual Rev. Gen.*, 19(1):253–272, 1985.

[9] Z. Wu, T. Jiang, and W. Su. Efficient computation of shortest absent words in a genomic sequence. *Inf. Proc. Let.*, 110(14):596–601, 2010.

Segmentation of the metacarpus, phalange and extensor tendon in musculoskeletal ultrasound images

Nelson Martins[12]
nelsonmartins89@gmail.com
Saad Malik[13]
engr.saadsultan@gmail.com
Diana Veiga[24]
dveiga@enermeter.pt
Miguel Coimbra[13]
mcoimbra@dcc.fc.up.pt
Manuel Ferreira[24]
mferreira@enermeter.pt

[1] Instituto de Telecomunicações
Porto, Portugal
[2] Enermeter, Sistemas de Medição, Lda
Braga, Portugal
[3] Faculdade de Ciências, Departamento de Ciências de Computadores
Universidade do Porto, Portugal
[4] University of Minho, Centro Algoritmi Azurém, 4800-058, Guimarães, Portugal

Abstract

In this work is presented a preliminary study concerning the automatic segmentation of anatomic structures in ultrasound images of the second metacarpophalangeal joint (MCPJ). These structures are known to be the first to be affected by rheumatic diseases. The objective is to create a system to identify these changes, in order to detect pathologies and do the follow up patients under treatment. Results proved that the automatic segmentation is possible, however more precise methods are needed to reduce errors.

1 Introduction

Rheumatic diseases are the main cause of long-term pain in the developed countries, which makes them a critical social, health and economical problem [3]. Due to its main advantages, ultrasounds are now being used in rheumatology to diagnosis and evaluate rheumatic diseases in early stages. The study of small joints using ultrasound imaging is accepted as a way to diagnosis and follow-up several rheumatic diseases [3]. To our knowledge there is no system to quantify changes in this images. A system of this nature could help rheumatologist in the follow up of patients under treatment, turning it more objective.

1.1 Second Metacarpofalangeal Joint - MCPJ

The aim of this work is to automatically identify the structures present in the MCPJ ultrasound images. In Figure 1 is shown the exam apparatus and the result image. The main structures present in these images are the metacarpus (Figure 1 - M), the phalange (Figure 1 - P) and the extensor tendon (Figure 1 - T). Given this, we divided the problem in two parts, bone segmentation and tendon segmentation. Next, the methods used to segment each structure are presented and after that the results and conclusions.

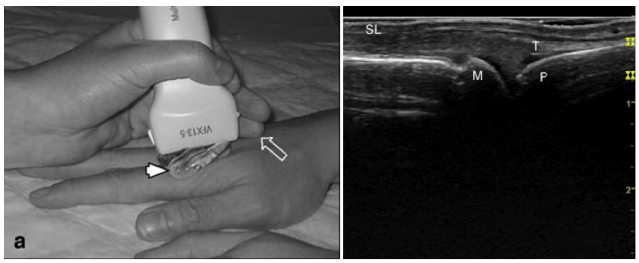


Figure 1: Ultrasound image of the second metacarpophalangeal joint. In the left is shown the probe position and in the right the resulting image. SL-Skin Line; T-Tendon; M-Metacarpus; P-Phalange. (Structures appear above the respective letter)

2 Proposed Work

2.1 Metacarpus and Phalangeal Bones

The metacarpus and phalange are two bones of the hand. In ultrasound images their appearance is similar, and that's the main reason to use the same segmentation method to identify them. Visually they are characterized as bright horizontal structures, due to high reflectance of the ultrasound waves. The technique used in the bone detection was based on active contours, more precisely, the technique described in [2], LAC - Local Active Contours. This method starts from a coarse segmentation and, iteratively, refine it until a more precise segmentation is achieved. Active contours methods are based in two energies, one referent to the image itself and the other referent to the expected shape of the object. These energies are known, respectively, as external and internal energy and both are responsible for the movement of the contours in each iteration, so that their sum is minimized. The external energy is normally associated to the image gradients and the internal energy associated with the line that defines the contours. In the LAC method, these energies are calculated locally, allowing a natural adaptation to non-uniformities of the background. This characteristic is useful because, in these images, the background is different along the image. For instance, the background in the joint capsule area is darker than the background of the peripheral zones, Figure 1.

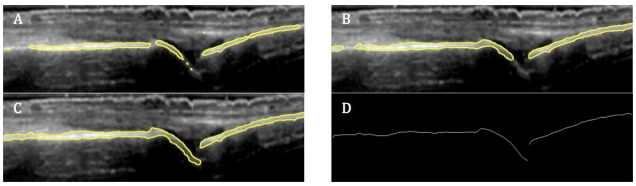


Figure 2: Metacarpus and Phalange segmentation steps. {A, B, C} - LAC iterations and D - Final result

The proposed method starts with the trimming of the bottom half of the image, followed by a downsampling by a factor of 2. This step reduces the amount of information to be processed without interfering with the overall segmentation performance. Next, a coarse segmentation is performed, we look vertically for horizontal bars with high values of intensity, and the one with higher intensity is set as foreground and the others as background. After, the LAC algorithm is used to improve the segmentation, Figure 2 - {A, B, C}. The external energy was calculated in relation to the image gradients, and the internal energy was associated to the first derivative of the line formed with the pixels of the contours. The parameters selected were the number of iterations, and the neighborhood radius of the local component, which were set, respectively, to 800 iterations and 5 pixels. The resulting mask is up-sampled to the original size and artifacts are removed (small structures). At last, the upper line is extracted from the resulting mask, Figure 2 - D. At this point, we expected to have 2 main lines, one representing the metacarpus and other the phalange. To separate them we only assume that the metacarpus is the one in the left and the phalange the one on the right. In some images the algorithm returns 3 structures, this is due to discontinuities in the bone,

as in the case of osteophytes. In these cases the structure in the middle is connected to the closest structure.

2.2 Extensor Tendon

The extensor tendon is a horizontal continuous dark line, bellow the skin line and above the metacarpus and phalange, Figure 1 - T. To identify the tendon was necessary to implement a preprocessing stage, due to low contrast. After a literature review, the phase symmetry was chosen [1]. This technique use Gabor filters to enhance valleys and hills like structures in different scales and orientations. The concept of phase symmetry method arrives from the fact that the valleys and hills have a symmetry line, which corresponds to their lower/higher points.

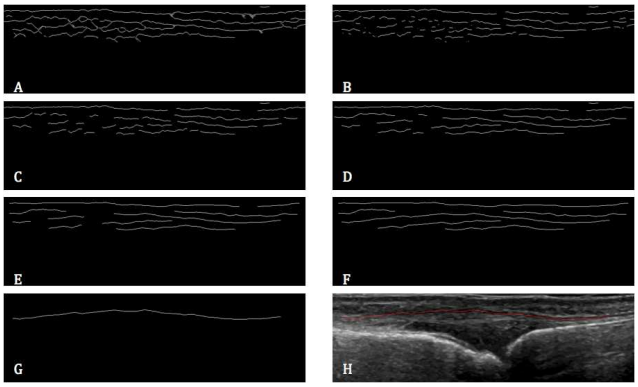


Figure 3: Diferent steps of the tendon segmentation

This characteristic is used to enhance the tendon, given its similarity to a valley. The enhanced image is then used to detect all horizontal lines. All these lines are considered candidates, figure 3 - A. At this point the number of candidates is still high, and with a lot of noise. To reduce the false candidates was used a filter and merge technique. First, lines with a slope higher than 30° are removed, figure 3 - B, then the small lines are also removed and, after that, the extremes of each line are connected, if they have other extreme close to it, figure 3 - C. The remaining lines are smoothed before another step of filtering and merging, figure 3 - {D, E, F}. From the resulting image, the two bigger lines are selected and the one below is considered to be the tendon figure 3 - G. This criteria was used because in most of the cases the superior line was the interface of the skin line with the ultrasound probe.

3 Results and Discussion

The proposed algorithms were tested on a set of 164 images, with ground truth validated by a rheumatologist. In table 1 the results obtained for each structure are shown. The metric used was the sensitivity and the specificity, to test the capacity to detect the structures and exclude outliers. This results were obtained by comparison between the manual and automatic segmentation, pixel by pixel and assuming only vertical correlations, for instance, if a given column has a pixel that belong to the manual and automatic segmentation, it is assumed to be a true positive.

Table 1: Sensitivity and Specificity obtained for each structure

	Metacarpus	Phalange	Tendon
Sensitivity	92.36	95.00	89.81
Specificity	90.90	92.26	50.70

Looking at the results is possible to see that the metacarpus and phalange results were similar. This result was expected because the method used was the same and their characteristics are similar. The results obtained for the tendon were inferior, with a specificity of 50%. This result is due to the fact that the algorithm used forces the existence of tendon along all image. This assumption is anatomically valid, but it is not always visible in the US image, because of poor positioning of the probe. One possible reason for the low specificity values in tendon segmentation is related with the ground truth definition, where only visible tendon was identified. Another important aspect to take into consideration is the average segmentation error. The measurement of the error was made vertically using the Euclidean distance. If this distance was smaller then a given error it was considered a true positive, otherwise it as considered a false negative. The error was plotted with the sensitivity, in order to have a better viewing of the confidence interval for each error.

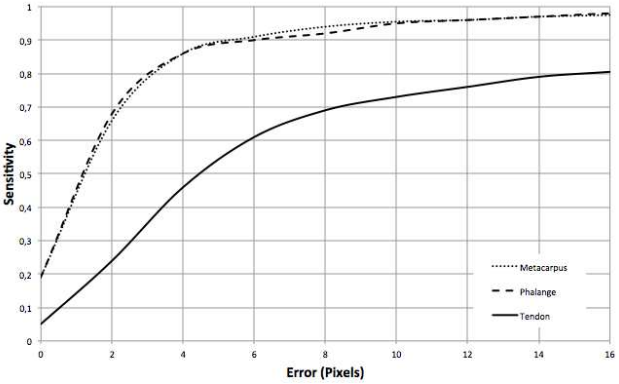


Figure 4: Segmentation error for each detected structure

Once again, the segmentation of the metacarpus and the phalange led to similar errors; around 80% of the structures were segmented with an error smaller than 3 pixels (considered an acceptable error, corresponding to 0.16mm). Regarding the tendon 73% of the structures were segmented with an error below 10 pixels. This result is outside the expected limits, so further work must be done.

4 Conclusions

A new technique for the segmentation of the metacarpus, phalange and extensor tendon in images of the second MCPJ was proposed. The results obtained proved that is possible to automatically identify these structures. Nevertheless, results need to be improved since they are still below the expected. This preliminary study can be used as a starting point to other works, because the tendon segmentation is still an open problem. The proposed phase symmetry preprocessing allowed the use of a simple segmentation approach that, despite its simplicity, proved to work in most cases. In the future a line fitting approach will be tested in the segmentation of the tendon, using the bone segmentation to extract statistic information about the possible position of the tendon (like minimum, mean and maximum distance from bone to tendon). This way we expect to reduce the area of interest and enhance the results.

5 Acknowledgments

This work is funded by Instituto de Telecomunicações in the scope of Project Rheumus (Projeto QREN no: 38505) and by the Fundação para a Ciência e Tecnologia (FCT) grant no: PD/BD/105761/2014 and in the scope of the Project HeartSafe (PTDC/EEI-PRO/2857/2012)

References

[1] Peter Kovesi et al. Symmetry and asymmetry from local phase. In *Tenth Australian joint conference on artificial intelligence*, volume 190. Citeseer, 1997.

[2] Shawn Lankton and Allen Tannenbaum. Localizing region-based active contours. *Image Processing, IEEE Transactions on*, 17(11): 2029–2039, 2008.

[3] Alexander K Scheel, Kay-Geert A Hermann, Elke Kahler, Daniel Pasewaldt, Jacqueline Fritz, Bernd Hamm, Edgar Brunner, Gerhard A Müller, Gerd R Burmester, and Marina Backhaus. A novel ultrasonographic synovitis scoring system suitable for analyzing finger joint inflammation in rheumatoid arthritis. *Arthritis & Rheumatism*, 52(3):733–743, 2005.

Perspectives for digitally-enabled engagement tools within prospective surveillance model for breast cancer survivors

João P. Monteiro
jpsm@ieee.org
Luís F. Teixeira
luisft@fe.up.pt
Hélder P. Oliveira
helder.f.oliveira@inesctec.pt

INESC TEC and Faculdade de Engenharia, Universidade do Porto
Campus da FEUP
Rua Dr. Roberto Frias
4200-465 Porto
Portugal

Abstract

Medicine has evolved substantially since Hippocrates. This change may be largely perceived throughout the availability of technical advances, but it also expresses a paradigm shift triggered not only by better comprehension of the diseases but also the perception we have of its impact.

A case in point is breast cancer, the most common cancer both in developed and developing regions, which treatment techniques result in several impairments, and, consequently, contribute to a decreased quality of life. While the assessment of the oncological outcome of the treatment can be easily objectively quantified, the same does not withstand for functional aspects closely related to quality of life. Though some methods for monitoring and assessing do exist, an integrated approach able to achieve early detection, promote risk-reduction and self-management, while engaging the patient in an appropriate follow-up strategy, is still missing.

The objective of this work is to identify potential contributions in need for the development of future breast cancer survivors (BCS) surveillance programs, trying to expound on the aspect of digitally enabled healthcare resources as engagement tools while seeking to contribute to the development of an effective personalized, prospective medicine program.

1 Introduction

Surgical treatment of breast cancer has been evolving. Obligatory mastectomy has been replaced by a well-established trend towards the adoption of breast-conserving approaches. The aim of breast cancer conserving treatment (BCCT) is to be as effective as mastectomy in terms of oncological outcome while promoting better cosmesis, better self-esteem and less psychological morbidity. This change is largely due to the availability of technical advances in terms of therapeutic methodologies, but it also expresses a paradigm shift triggered by better comprehension of both the disease and the perception we have of its impact [20].

As BCS are living longer, the adverse effects resulting from the cancer treatment are more frequent. Upper body morbidity (UBM) (e.g. decreased range of motion, muscle strength, pain and lymphedema) are among the most prevalent side effects ([14], [17]). While the assessment of the oncological outcome of the breast cancer treatment (BCT) can be easily objectively quantified by disease-free and overall survival rates, the same does not withstand for functional aspects closely related to quality of life (QOL). Assessment of BCS symptoms and health-related quality of life (HRQOL) outcomes are usually made using patient-reported outcome (PRO) questionnaires, that quantify significant outcome variables from the patient's perspective ([4], [16]). However, there are no uniform standardized objective assessment criteria for assessing the upper body function (UBF) in BCS [14]. Furthermore, it has yet to be determined which clinicians (breast cancer specialists, family practitioners, rehabilitative clinicians) are formally responsible for the diagnosis, treatment and management of survivors' UBM related care needs [17], thus enhancing the need of a proper UBF assessment tool that enables a more active role of the patient.

Hereupon, this work presents a brief contextualization of a framework outline for the future development of digitally-enabled engagement tools within prospective surveillance model for BCS.

2 Background and Related Work

2.1 Lymphedema and UBF Impairments

Regarding lymphedema alone, it has been estimated that over 1 million BCS in the United States [14] and 10 million women worldwide [17] may meet the criteria for breast cancer-related lymphedema (BCRL). Lymphedema is a swelling condition, resulting from lymphatic ablation commonly associated with BCT. Women who have undergone BCT are at risk of developing BCRL during their lifetimes, which impacts on different dimensions of a woman's QOL, including her physical, psychological, and emotional well-being [9]. Clinical assessment of the condition is usually performed by evaluation of the difference in volume between the operated side and the other, or, against a pre-operative measurement. Although there is a lack of consensus for standardized protocols in regard to measurement techniques [14], objective methods include bioimpedance spectroscopy, arm circumferences, water displacement or lymphoscintigraphy. More recently a Kinect-based system was proposed and evaluated for the purpose of estimating upper-limb volume [3].

It is, however, possible to identify other aspects of interest to further evaluate function that may not be necessary related with BCRL, that result in limitations in activities of daily living. Arm/shoulder mobility, usually assessed by goniometer-based measurements of flexion, is an objective measure of UBF that has been used in the breast cancer rehabilitation. More recently, several studies proposed the use of Kinect for UBF assessment, trough estimation of reachable workspace based on hand and shoulder trajectories point clouds ([7], [12]).

2.2 Serious games as digitally-enabled engagement tools

Besides UBF assessment itself, recently proposed prospective surveillance model for BCS [19] highlight the importance of monitoring for functional and physical impairment commonly associated with BCT. Such concern has been recurrently reiterated ([14], [17]). It can also be noted that since adverse effects can occur years after BCT, long-term follow-up strategies and objective widely available assessment methods are presented to be in need to materialization of such surveillance.

In this sense, growing trends of the quantified-self movement, personal health records tools dissemination and interactive video games that combine physical exercise with game-play and have a primary purpose other than entertainment [1] present themselves as currently active research lines with great potential [5]. While there are several games that include serious topics (e.g. [13]), the inclusion of serious game elements is not yet enough to induce learning or real-world action [18]. Recently proposed cognitive behavioural game design (CBGD) framework suggests that such real-world behaviour change in the player should be accomplished trough incorporated into game design the use of elements in social cognitive theory (SCT) and multiple intelligences (Mis) theory in an enjoyable way via the use of in-game libraries.

More broadly, engaging patients in their healthcare can be recognized as a growing trend that also reflects the increasing world population with chronic disease ([15]). However, although the concept of patient engagement being increasingly presented as yielding great potential to improve healthcare, many gaps can be identified for the latter to be achieved, starting by defining the concept of *patient engagement* itself [6]. This term may not only be used to refer to different types of interactions between patients and health care systems but also it is possible to find different terms with overlapping definition (i.e. patient activation, patient involvement, patient participation, patient adherence, patient empowerment and patient compliance). As presented in [6] a systemic conceptualization of the patients' role in healthcare is presented. In the aforementioned work of Graffigna *et al.* [6] it is also described the Patient Health Engagement Scale (PHE-scale), a measure of patient engagement grounded in psychometric methods. Patient Activation Measure (PAM) [8] should also be mentioned as a concurrent measure, although the concept of *activation* is more limited to the situation of a doctor-patient consultation.

3 Serious Games for Breast Cancer Survivors

Regarding lymphedema and UBF impairments, besides no uniform standardized objective assessment criteria has found, it was shown that some patients develop symptoms of BCRL without objective changes in arm circumference, indicating that clinical measurements may underestimate its incidence and impact [16]. Other than that, the correlation of such limitations with HRQOL has not been systematically evaluated [4]. Furthermore, current models of the shoulder and upper limb still present limitations as tools for robust UBF analysis with clinical relevance [2].

Although engaging patients in their healthcare can be recognized as a current trend, methods for clinical and at home use to eliminate biases and recall inaccuracies from self-report data, as well as achieve early detection, promote risk-reduction and self-management procedures are still missing [11]. Although multiple works have been recently published, examination of the efficacy of intervention to the individual patient needs is still missing. Overall, despite engagement being considered a valuable resource, research on patient engagement technologies regarding impact on health outcomes has been limited [15].

In line with the aforementioned discussion, this work proposes that an integrated approach for BCS physical impairments early detection, integrated evaluation of surgical outcome and self-management promotion should be comprised of two following main elements:

3.1 Objective evaluation of functional outcome after BCT

Within this first part, the purpose is at providing measures of the outcome after particular treatment options to the clinicians, in order to them not only to perform a better informed decision in future treatments, but also enable improved BCS follow-up. This is to be accomplished through the over time assessment of such quantities as circumferential arm measures, range of motion but mainly the apprehension of the prescribed exercise performance. Highlight here the challenges related with quantification of current rehabilitation procedures and identification of functional impairments that precede more advanced states of UBM.

In that manner, recent computer vision and machine learning methods present great potential, given the availability of off-the-shelf markerless sensors that full-fill requirements of low cost, easy and widespread use [10]. Nonetheless, given the necessity of high precision of measurements of not only upper body joints position and orientation but also objects with changing appearance in cluttered and occluded scenes, the tasks of model and assess the upper extremity functional capability in respect to BCS in a unconstrained manner, should be regarded as an open challenge.

3.2 Serious games tailored for BCS

In order to promote early detection of UBM, education on risk-reduction and self-management habits through the application of serious games, as it was previously introduced, it can be presented as in need the study and comparison of current approaches on interface and game design, such us the use of communicating and/or reinforcing knowledge, use of recommender systems given user in game assessment and the exploration of interface embodiment alternatives. This should allow future BDS surveillance programs to be comprehensive, as games should besides deduce from the participants' behaviour objective UBF measures with interest to physicians, enable BCS independent and playful use of engaging and exercise promoting activities. Ultimately, the evaluation of the alternative game solutions should be performed with recourse to a patient health engagement scale (i.e. PHE-scale, PAM, ...), as well as further demographic and clinical variables, in order to describe socio-demographic and clinical characteristics of the patients.

4 Conclusion

Common breast cancer treatment techniques result in several impairments in women's upper-body function, and, consequently, contribute to a decreased quality of life. Methods for monitoring and assess are not only missing as it possible to recognize the potential of interdisciplinary collaboration to the materialization of an effective personalized medicine program in breast cancer-related upper body morbidity assessment and prevention.

Acknowledgment

This work is financed by the FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013.

References

[1] F Bellotti, B Kapralos, K Lee, P Moreno-Ger, and R Berta. Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction*, 2013:1–11, 2013.

[2] B Bolsterlee, D H Veeger, and E K Chadwick. Clinical applications of musculoskeletal modelling for the shoulder and upper limb. *Med Biol Eng Comput*, 51(9):953–963, 2013.

[3] R Buffa, E Mereu, P Lussu, V Succa, T Pisanu, F Buffa, and E Marini. A new, effective and low-cost three-dimensional approach for the estimation of upper-limb volume. *Sensors*, 15(6):12342–12357, 2015.

[4] K L Campbell, A L Pusic, D S Zucker, M L McNeely, J M Binkley, A L Cheville, and K J Harwood. A prospective model of care for breast cancer rehabilitation. *Cancer*, 118(S8):2300–2311, 2012.

[5] Moheb Costandi. Rehabilitation: Machine recovery. *Nature*, 510(7506):S8–S9, 2014.

[6] G Graffigna, S Barello, A Bonanomi, and E Lozza. Measuring patient engagement: development and psychometric properties of the patient health engagement (PHE) scale. *Front Psychol*, 6, 2015.

[7] N Hawi, E Liodakis, D Musolli, E Suero, T Stuebig, L Claassen, C Kleiner, C Krettek, V Ahlers, and M Citak. Range of motion assessment of the shoulder and elbow joints using a motion sensing input device. *Technol Health Care*, 22(2):289–295, 2014.

[8] Judith H. Hibbard, Jean Stockard, Eldon R. Mahoney, and Martin Tusler. Development of the patient activation measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Serv Res*, 39(4p1): 1005–1026, 2004.

[9] J T Hidding, C H Beurskens, P J van der Wees, H W van Laarhoven, and M W van der Sanden. Treatment related impairments in arm and shoulder in patients with breast cancer: A systematic review. *PLoS ONE*, 9(5):e96748, 2014.

[10] H M Hondori and M Khademi. A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *J Med Eng*, 2014: 1–16, 2014.

[11] C A Hudis and L Jones. Promoting exercise after a cancer diagnosis: easier said than done. *Br J Cancer*, 110(4):829, 2014.

[12] L P Lowes, L N Alfano, B A Yetter, L Worthen-Chaudhari, W Hinchman, J Savage, P Samona, K M Flanigan, and J R Mendell. Proof of concept of the ability of the kinect to quantify upper extremity function in dystrophinopathy. *PLoS Curr*, 2013.

[13] J Marques, A Vasconcelos, and L Teixeira. Senior-driven design and development of tablet-based cognitive games. *Stud Health Technol Inform*, 189: 133–138, 2013.

[14] P Ostby, J Armer, P Dale, M Van Loo, C Wilbanks, and B Stewart. Surveillance recommendations in reducing risk of and optimally managing breast cancer-related lymphedema. *J Pers Med*, 4(3):424–447, 2014.

[15] J E Prey, J Woollen, L Wilcox, A D Sackeim, G Hripcsak, S Bakken, S Restaino, S Feiner, and D K Vawdrey. Patient engagement in the inpatient setting: a systematic review. *J Am Med Inform Assoc*, 21(4):742–750, 2013.

[16] A L Pusic, Y Cemal, C Albornoz, A Klassen, S Cano, I Sulimanoff, M Hernandez, M Massey, P Cordeiro, M Morrow, and B Mehrara. Quality of life among breast cancer patients with lymphedema: a systematic review of patient-reported outcome instruments and outcomes. *J Cancer Surviv*, 7(1): 83–92, 2012.

[17] R Shaw and R Thomas. The information needs and media preferences of canadian cancer specialists regarding bct related arm morbidity. *Eur J Cancer Care*, 23(1):98–110, 2013.

[18] K Starks. Cognitive behavioral game design: a unified model for designing serious games. *Front Psychol*, 5, 2014.

[19] M D Stubblefield, M L McNeely, C M Alfano, and D K Mayer. A prospective surveillance model for physical rehabilitation of women with breast cancer. *Cancer*, 118(S8):2250–2260, 2012.

[20] R Thomas and R Hamilton. Illustrating the (in)visible: Understanding the impact of loss in adults living with secondary lymphedema after cancer. *Int J Qual Stud Health Well-being*, 9(0), 2014.

Pedestrian Detection using Spatial Pyramid Pooling in Deep Convolutional Networks

Miguel Farrajota
mafarrajota@ualg.pt
J.M.F. Rodrigues
jrodrig@ualg.pt
J.M.H. du Buf
dubuf@ualg.pt

Vision Laboratory, Inst. for Systems and Robotics,
LARSyS, University of the Algarve
Campus de Gambelas, 8005-139 Faro, Portugal

Abstract

Pedestrian detection and tracking remains a popular issue in computer vision, spawning many applications in robotics, surveillance and security, biometrics and human-computer interaction. We present an improvement to pedestrian detection, based on deep convolutional networks, in terms of speed performance. By employing spatial pyramid pooling (SPP) instead of usual max pooling over the last convolution layer, images of arbitrary sizes can be used for model training and evaluation, also enabling full pedestrian detection in a single pass.

1 Introduction

Pedestrian detection is a very challenging task due to the large variability caused by different clothing and poses, abundant partial occlusions, complex/cluttered backgrounds and frequent changes in illumination. In recent years, considerable progress in the development of approaches and applications has been obtained concerning object detection and class-specific segmentation in tracking scenarios, pedestrian detection being of particular interest [3]. All existing state-of-the-art methods use a combination of bio-inspired [6] or hand-crafted features such as HoG [1], Integral Channel Features [2] and other variations [7] and combinations [12], along with a trainable classifier such as boosted classifiers [2], SVM [7] or random forests [4]. Although low-level features can be designed by hand with good success, mid-level features composed by combinations of low-level features are difficult to engineer without employing some sort of learning procedure. Multi-stage classifiers that learn hierarchies of features tuned to specific objects can be trained end-to-end with little prior knowledge. Convolutional Neural Networks (ConvNets/CNNs) [9] are examples of such hierarchical systems which are trained in a supervised manner.

The contribution of this paper consists of combining simple region proposals produced by a sliding window technique over convolution features' pool (last convolution layer) with spatial pyramid pooling. Computations are reduced and real-time detection of pedestrians can be achieved in order to be competitive with the state of the art in both accuracy and speed. The system consists of a convolutional network for feature learning and classification, combined with region proposals obtained by a sliding window method with varying sizes, and spatial pyramid pooling composed of several pooling regions (1×1 to 3×3) over the last convolution layer. The advantage is that arbitrarily sized images can be processed in a single forward propagation through the network, reducing feature computations over scales and optimizing filter convolutions [8]. Feature learning and model fine-tuning was done using the INRIA dataset [1].

2 Pedestrian detection

In this work, pedestrians are detected using a sliding window over the image similar to many popular and state-of-the-art methods based on a CNN, e.g. [9]. The method consists of two steps: (a) convolve only once the entire image with the feature extraction layers of the CNN, thus avoiding expensive computations at multiple scales, and (b) slide the classifier over the resulting feature maps obtained in (a) with varying window sizes in order to detect pedestrians of different sizes.

2.1 Network design and training

We use standard fully supervised convnet models [9] for feature extraction and category (pedestrian/background) detection. The model is divided into two modules, namely: (a) *feature extraction* and (b) *classification*. In (a), each layer consists generically of (i) convolution of the previous layer output (or, in the case of the 1st layer, the input image) with a set of learned filter kernels; (ii) passing the responses through a rectified linear unit ($\mathrm{ReLU}(x) = \max(x,0)$), and (iii) max pooling over local neighborhoods. More specifically, concerning the convolution layers, in the first one we apply 32 filters with size a of 7×7 pixels and stride of 2 pixels in x and y, zero-padding the image with borders of 2 pixels. In the next two convolution layers we apply 64 and 128 filters, respectively, with a filter size of 3×3 and stride 1. All spatial max pooling layers pool over regions of 2×2 with strides of 2.

(b) For *Classification*, the top two layers of the network are conventional fully-connected (FC) networks and the final layer is a softmax classifier. The first FC net has 512 hidden units with 50% dropout and the ReLU function. It is connected to the second FC net with 2 outputs. In the feature extraction section, we replace the last max pooling layer by a spatial pyramid pooling one, with pooling windows of size 1×1, 2×2 and 3×3. The convolution layers accept arbitrary input sizes, but they produce outputs of variable sizes. Since the classifiers (SVM/softmax) and fully-connected layers require fixed-length vectors, with spatial pyramid pooling the input image can be of any size. This allows inputs to be resized to any scale and to apply the same deep network. Also, after the network is trained, the fully connected layers are replaced by convolution layers [11] to speed up computations when evaluating the model.

We train the network on the INRIA dataset (2416 positive and 1218 negative samples). Since this is a relatively small dataset, we employ a data augmentation strategy to reduce over-fitting and to increase the accuracy of the classifier. We use mini-batches of 128 samples, with positive and negative samples randomly selected with the same probability (50% chance) over 1000 iterations in each epoch. We train the model during 60 epochs with randomly initialized weights. These weights are updated using stochastic gradient descent with a momentum of 0.9 and a learning rate of 10^{-2}, a weight decay of 5×10^{-4} over the first 30 epochs, and decreasing the learning rate to 10^{-3} and 10^{-4} after 20 and 40 epochs. During training, our model receives inputs of variable sizes. Positive images vary in size from 32×64 pixels to 192×384 pixels. Negative images vary from 648×486 pixels down 64×64 pixels where a single random crop is extracted. For positive and negative samples, we crop a single region ranging from 32×64 up to 192×384 pixels, depending on the image size. For positive images, we only crop regions if at least 70% of the pedestrian's ground truth bounding box is contained inside the region. Scale sampling (jittering) is performed on the training data prior to cropping for positive and negative sample images, where sizes range from $\min(w,h) \geq 32$ and $\max(w,h) \leq 512$, w being the width and h the height. We perform color casting to alter the intensities of the RGB channels in training data. Specifically, for each image, we randomly change the R, G and B values up to $\pm 20\%$ with 50% probability.

2.2 Detection

The trained network takes as input a single-scale image and region (R) proposals which are obtained after applying the convolution layers over the image and sliding the spatial pyramid layer over the last convolution layer. To this effect, we use windows of varying sizes starting with $R_w = 8$ and $R_h = 16$ grid sizes, with steps of 1 grid pixel (this corresponds to steps of 4 pixels in the input image), with a window growth factor of 20% if the window's size is smaller than the image itself, i.e., $\min(R_w, R_h) < \min(w,h)$. For a 640×480 image, the typical window size is 10^4 pixels. For each test region R, ta class posterior probability distribution is computed, all regions classified as background are removes, and then non-maximum suppression is performed to separate the strongest from the weakest detection (at least 0.65% overlap). The process achieves around 3 fps in case of images having a maximum dimension ≤ 600 pixels, using a Nvidia Titan GPU.

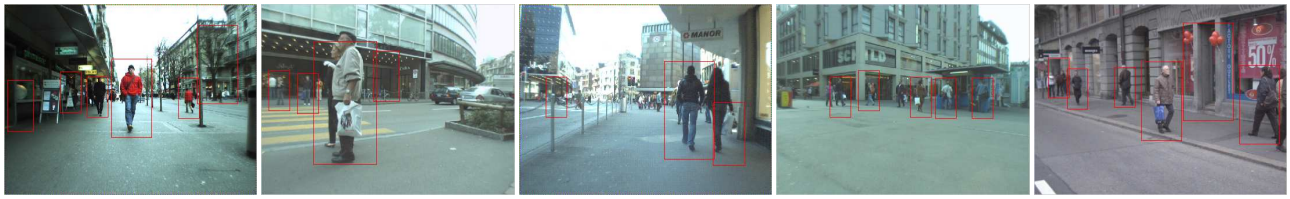


Figure 1: Detection results on the ETH dataset [5].

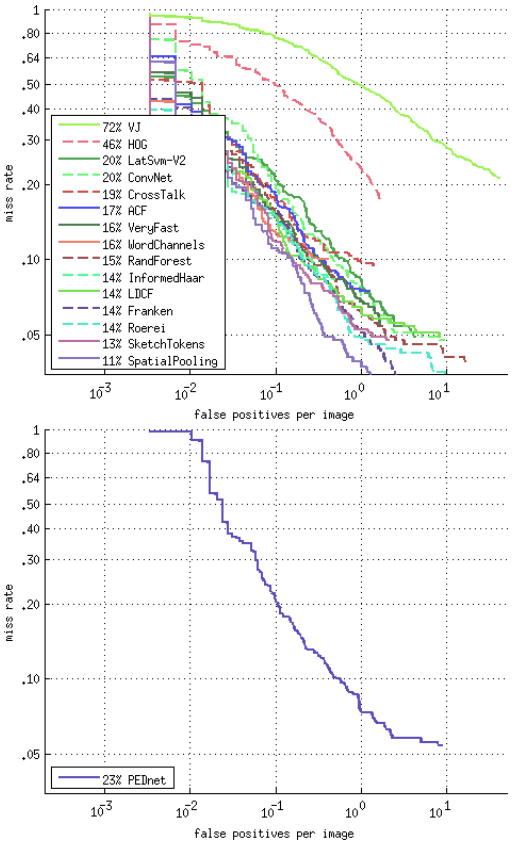


Figure 2: Top, INRIA dataset for large pedestrians benchmark with available top-performing methods [3], and (bottom) our method's. Lower curves indicate better performance.

3 Discussion and results

This paper presented an optimization strategy for speed on pedestrian detection using CNN networks. By computing features on a single scale, the computational cost of processing features on multiple scales of the same image is greatly reduced. Then, by using a spatial pyramid pooling layer in a sliding window fashion over the final feature map, images of arbitrary sizes can be processed in a single forward pass. This pooling layer allows to pool features from differently sized regions and to feed them into the classifier which requires fixed-sized feature arrays as input. To complement the reduction in computation, we only use a single scale for processing, and fully connected layers were replaced by convolution ones which offer more efficient computations on GPUs. These improvements significantly help reducing the overall speed constraints of convnet's applications, allowing tasks such as pedestrian detection to be performed in practical time (3 fps on average). Although there is still room for improvement, the versatility of deep nets for recognition tasks compared to specially tailored systems built for specific tasks outweighs its weaknesses concerning speed and accuracy when more complex scenarios (surveillance, tracking, etc.) are involved.

Results of the method demonstrate the applicability and usefulness of detector. Figure 1 shows some results on the ETH dataset [5]. Figure 2 shows two DET curve plots: the overall performance of our method (bottom) against other top-performing methods available online [3] (top).

For future work, it is intended to further refine the detection method by experimenting with multi-stage features (features from previous conv layers) and by adding bounding box prediction to the classification. This will allow to use bigger steps and window sizes when sliding the spatial pyramid pool layer over the convolution features, further reducing the computational cost. Tests with pre-trained networks such as alexnet [9] and VGG [10] for feature extraction will be performed in order to evaluate the trade-off between speed and accuracy. Also, a more demanding dataset such as Caltech [3] will be explored.

Acknowledgments

This work was supported by the FCT project LARSyS: UID/EEA/50009/2013 and FCT PhD grant to author MF (SFRH/BD/79812/2011).

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, I:886–893, 2005. ISSN 1063-6919. doi: 10.1109/CVPR.2005.177.

[2] P. Dollár, Z. Tu, P. Perona, and S. Belongie. Integral Channel Features. *BMVC 2009 London Engl.*, pages 1–11, 2009. doi: 10.5244/C.23.91.

[3] P Dollár, C Wojek, C Schiele, and P Perona. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell.*, 99(4):743–761, 2011.

[4] P. Dollár, R. Appel, and W. Kienzle. Crosstalk cascades for frame-rate pedestrian detection. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, volume 7573 LNCS, pages 645–659, 2012. ISBN 9783642337086. doi: 10.1007/978-3-642-33709-3_46.

[5] A. Ess, B. Leibe, K. Schindler, and L. Van Gool. A mobile vision system for robust multi-person tracking. In *Comput. Vis. Pattern Recognition, 2008. CVPR 2008. IEEE Conf.*, pages 1–8, 2008. ISBN 1424422426. doi: 10.1109/CVPR.2008.4587581.

[6] J.M.H. Farrajota, M., Rodrigues, J.M.F., du Buf. Bio-Inspired Pedestrian Detection and Tracking. *3rd Int. Conf. Adv. Bio-Informatics, Bio-Technology Environ. Eng. Birmingham, UK, 26-27 May*, pages 28–33, 2015. doi: 10.15224/978-1-63248-060-6-07.

[7] P. F Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–20, 2009. ISSN 1939-3539. doi: 10.1109/TPAMI.2009.167.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv Prepr. arXiv ...*, cs.CV:1–14, 2014. ISSN 0162-8828. doi: 10.1109/TPAMI.2015.2389824.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.*, pages 1–9, 2012. ISSN 10495258.

[10] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recoginition. *Intl. Conf. Learn. Represent.*, pages 1–14, 2015.

[11] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for Simplicity: The All Convolutional Net. *Iclr*, pages 1–14, 2015.

[12] S. Walk, N. Majer, K. Schindler, and B. Schiele. New features and insights for pedestrian detection. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 1030–1037, 2010. ISBN 9781424469840. doi: 10.1109/CVPR.2010.5540102.

Generating datasets with drift

Joana Costa[12]
joana.costa@ipleria.pt,joanamc@dei.uc.pt
Catarina Silva[12]
catarina@ipleiria.pt,catarina@dei.uc.pt
Mário Antunes[13]
mario.antunes@ipleiria.pt,mantunes@dcc.fc.up.pt
Bernardete Ribeiro[2]
bribeiro@dei.uc.pt

[1] School of Technology and Management
Polytechnic Institute of Leiria, Portugal
[2] CISUC - Department of Informatics Engineering
University of Coimbra, Portugal
[3] Center for Research in Advanced Computing Systems
INESC-TEC, University of Porto, Portugal

Abstract

Modern challenges in machine learning include non-stationary environments. Due to their dynamic nature, learning in these environments is not an easy task, as models have to deal both with continuous learning process and also with the acquisition of new concepts. Different types of drift can occur, as concepts can appear and disappear with different patterns, namely sudden, reoccurring, incremental or gradual. Besides striving to propose new techniques to learn in the presence of drift, researchers aim to find appropriate benchmarks to non-stationary scenarios.

In this paper we propose DOTS, a drift oriented tool system, whose main goals are to define and generate datasets with drift for text classification problems. DOTS tries to fill an existing gap in machine learning research for text applications, by making possible to generate benchmark datasets with thoroughly controlled drift characteristics. Because of its ability to export in multiple formats, it can be widely used and in conjunction with well-known classifiers and applications, like SVM Light or WEKA. We will also present a Twitter case study to validate the usefulness of DOTS in a real-word problem scenario.

1 Introduction

Non-stationary environments are characterized by incremental data gathering where the underlying data distribution changes over time without an explicit and known pattern. The enormous growth of the computational power in recent years, along with the popularization of social networks and mobile devices, created a deluge of data and demanded new approaches, as all this information needs to be acquired and treated almost in real-time, as it evolves faster than we were used. There are nowadays multiple real-world applications where this is a given, like image or speak recognition, fraud detection or mining in social networks.

Learning in the presence of drift is not an easy task and requires a distinctive approach. One of the major challenges is posed due to the multitude of drift patterns, and the inability to perceive a priori which ones might be present. Zliobaite[1] identified four drift patterns, as can be seen in Fig. 1: sudden; gradual; incremental; and reoccurring.

- **Sudden drift** occurs when in a given moment a concept appears or disappears in an abrupt way. The speed of the drift is therefore high.
- **Gradual drift** occurs when the probability of a given context to be associated with a concept decreases during a certain period of time, but also the probability to be associated with another context increases proportionally.
- **Incremental drift**, sometimes considered a subgroup of gradual drift, can be considered differently because the change between the two concepts is very slow and only perceived when looking to what is occurring during a larger period of time.
- **Reoccurring drift** occurs when previously active concept reappears after a certain period of time. It is noteworthy to refer that the seasonality of the change must be previously unknown, otherwise the core assumption of the uncertainty about the future would be compromised.

Several recent approaches [2-4] try to deal with such challenges, namely by proposing dynamic techniques that try to detect, or deal, with drifts in different scenarios. But research challenges in learning in non-stationary environments are not only in proposing and deploying learning models able to learn in the presence of different drift patterns, but also to find suited benchmarks to test and compare the proposed approaches. To tackle this issue, we propose the Drift Oriented Tool System (DOTS). DOTS is a drift oriented framework able to define and generate text-based datasets with drift, that can then be used to evaluate and validate learning strategies for dynamic environments.

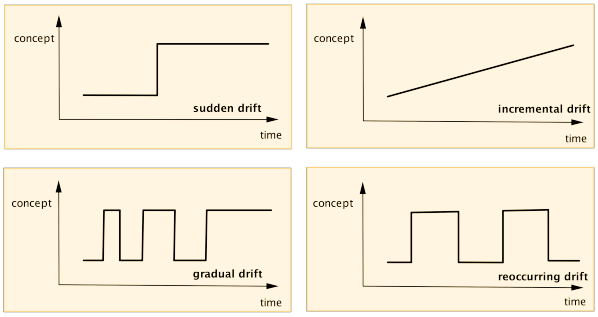


Figure 1: Types of drift.

2 Drift Oriented Tool System

DOTS is a drift oriented framework developed to dynamically create textual datasets with drift. It is a simple and easy to use freeware application with a friendly interface as shown in Fig. 2. It can be download at `http://dotspt.sourceforge.net/`.

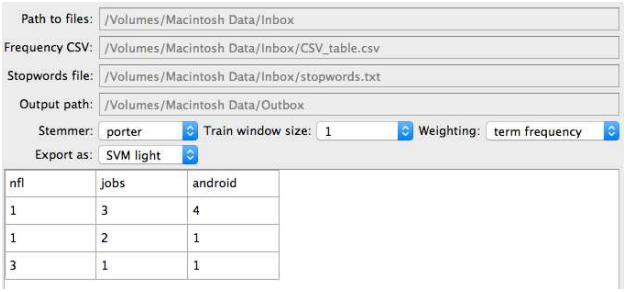


Figure 2: DOTS interface.

The main purpose of DOTS framework is to represent drift patterns in a text-based dataset. Therefore, the input of the DOTS framework is two fold and composed by text documents and a frequency table, as depicted in Fig. 3. Each text document file represents the documents of the same class and the frequency table is used to define the drift patterns of the scenario. A major characteristic of DOTS is the possibility of defining the time window where each document appears, being thus possible to define time drifts. This is done by the frequency table that is a mandatory input of the DOTS framework (see Fig. 3).

The frequency table must be in the CSV format and each row corresponds to a time instance. It is not important if a time instance represents a minute, an hour, or a day, but it is assumed that all of them correspond to the same amount of time. The first row contains the name of the class, and each of the above cells contain the number of documents of that

given class in the subsequent time windows. Fig. 2 depicts a task being added to the framework, in which the first time-window has one, three and four documents of the classes `nfl`, `jobs` and `android` respectively. The DOTS framework represents each document in a vector space

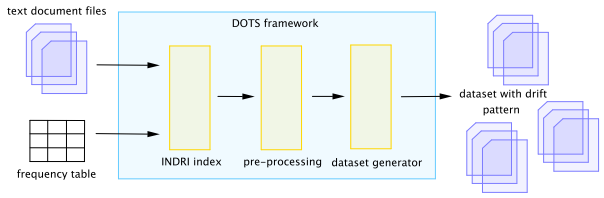


Figure 3: The DOTS framework.

model, also known as *bag of words*, a commonly used text document representation. Two problems arise when using the *bag of words*: high-dimensionality feature space and overfitting. High dimensional space can cause computational problems and overfitting can easily prevent the classifier from generalizing and thus endagered the prediction process. To tackle both problems pre-processing methods were also integrated in the DOTS framework and are the second phase of the DOTS processing. In order to use stopword removal a text file containing stopwords must be defined. These words, considered non-informative words, will not be included in documents representation. Besides stopwords removal, DOTS also allows stemming. This pre-processing method consists in removing case and inflection information of a word, reducing it to the word stem. Stemming does not alter significantly the information included, but it does avoid feature expansion. We have included two important stemming algorithms: the Porter algorithm[5] and Krovetz algorithm[6].

It is also possible to define the weighting scheme used to represent each word of a document, that is the weight of each feature of a document. Two weighting schemes were defined, namely term frequency (*tf*) and term frequency-inverse document frequency, commonly known as *tf-idf*. Considering the defined input, DOTS will create a word index, the INDRI index, provided by INDRI API, from the Lemur Project[1]. By outputting an INDRI index, DOTS provides all the features presented by the INDRI project, a powerful query interface, that provides state-of-the-art text search, field retrieval and text annotation.

It is also possible to define multiple training window sizes and multiple export file formats. The training window size will define in each time-window how many previous time-windows will be considered, as this is important to test learning models with memory capabilities. For instance, to perceive for how long it is relevant to keep previously gathered information and how that can affect the learning model capabilities, which can be seen as the importance of previously seen examples in future classifications. By exporting in multiple file formats, DOTS allows for creating datasets that can be used in different classification frameworks, like `SVM Light` and `Weka`. Three output formats were implemented: Comma-Separated Values (CSV) file format, the Attribute-Relation File Format (ARFF) used in the widely used `WEKA` software, and the `SVM Light` file format.

As it is often relevant to define various testing scenarios, DOTS permits adding tasks using INI files. These are structured files with "`key=value`" pairs, that contain the definition of multiple tasks. By using an INI file as input, users are able to define more than one task at a single time, thus optimizing the time spent on task setup. A complete tutorial can be download at `http://dotspt.sourceforge.net/`.

3 Case study: Twitter stream

To validate the usefulness of DOTS in a real-world problem, we will present a Twitter stream case study. It constitutes a paradigmatic example of a text-based scenario where drift phenomena occur commonly. *Twitter* is a micro-blogging service where users post text-based messages up to 140 characters, also known as *tweets*. *Twitter* is also responsible for the popularization of the concept of *hashtag*. An *hashtag* is a single word started by the symbol "#" that is used to classify the message content and to improve search capabilities. We have used a classification strategy previously introduced in [7], where the Twitter message *hashtag* is used to label the content of the message.

[1] `http://www.lemurproject.org/`

In this particular case, DOTS is used to create datasets able to test multiple learning scenarios. DOTS receives a document set for each class of tweets containing the same hashtag. A CSV table with different drift patterns was also defined, reproducing artificial drifts, like sudden, gradual, incremental, reoccuring and normal. As an example, a sudden drift might be represented by tweets from a given hashtag that in a given temporal moment start to appear with a significant frequency. Each tweet was represented by DOTS as a vector space model and pre-processing methods were applied, like stopword removal and stemming. We have exported, for our convenience, in SVMLight format, as is required to use Support Vector Machines (SVM) as the learning model of this case study. Two different weighting scheme in document representation were tested, term-frequency and *tf-idf*. Table 1 presents the obtained results using both weighting schemes. To evaluate the possible outcomes of the classification, we used the van Rijsbergen F_β measure: $F_\beta = \frac{(\beta^2+1)P \times R}{\beta^2 P + R}$ with $\beta = 1$. The obtained results showed that using term-frequency to represent document features improves the overall classification performance even in the presence of different drift patterns.

Drift	Hashtag	F_1 using tf	F_1 using tf-idf
Sudden	#syrisa	79.37%	75.32%
Gradual	#airasia	57.88%	56.08%
Incremental	#isis	83.49%	81.75%
Reoccurring	#android	60.66%	59.49%
Normal	#sex	74.88%	74.56%

Table 1: Performance measure for the results obtained with Twitter stream.

4 Conclusions and Future Work

In this paper we have presented the Drift Oriented Tool System. DOTS is designed for text-based problems and can produce datasets with multiple drift patterns and in multiple file formats, which is of major importance as it can be used with different classification frameworks. It aims to fulfil the existing gap in machine learning of tools able to reproduce benchmarks in dynamic environments.

We have also presented a case study based in a Twitter scenario, to validate the usefulness of using DOTS in real-world problems. We have generated multiple datasets and have tested different classification strategies, to define the best characteristic of a learning model in this particular scenario. All of these demonstrate the convenience of DOTS to create text-based datasets with drift and thus be used to evaluate and validate learning strategies in dynamic environments.

Regarding future work we will look at including new features in the DOTS framework, namely, the possibility of noise addition to the datasets and new file exporting types.

Acknowledgment

We acknowledge the iCIS project (CENTRO-07-ST24-FEDER - 107002003).

References

[1] Indre Zliobaite. Learning under Concept Drift: an Overview. Tech. Report, Vilnius University, Faculty of Mathematics and Informatic, 2010.

[2] Joana Costa, Catarina Silva, Mário Antunes, and Bernardete Ribeiro. Concept Drift Awareness in Twitter Streams. In *Proc. of the 13th International Conference on Machine Learning and Applications*, pages 294-299, 2014.

[3] D. Mejri, R. Khanchel, and M. Limam. An Ensemble Method for Concept Drift in Nonstationary Environment. In *Journal of Statistical Computation and Simulation, vol. 83, no. 6*, pages 1115-1128, 2013.

[4] G. Ditzler and R. Polikar. Incremental Learning of Concept Drift from Streaming Imbalanced Data. In *IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 10*, pages 2283-2301, 2013.

[5] P. Willett. The Porter Stemming Algorithm: Then and Now. In *Program, vol.40, no. 3*, pages 219-223, 2006.

[6] R. Krovetz. Viewing morphology as an Inference Process. In *Proc. of the 16th Annual International ACM SIGIR conference on Research and development in information retrieval*, pages 191-202, 1993.

[7] Joana Costa, Catarina Silva, Mário Antunes, and Bernardete Ribeiro. Defining Semantic Meta-Hashtags for Twitter Classification. In *Proc. of the 11th International Conference on Adaptive and Natural Computing Algorithms*, pages 226-235, 2013.

Neural Network Classifier for Gastroenterology Images - Paving the way for a MultiResolution approach

Vasco M.C. Ferreira
vascoferreira1@gmail.com
Miguel Tavares Coimbra
mcoimbra@dcc.fc.up.pt

Instituto de Telecomunicações
Faculdade de Ciências da Universidade do Porto
Portugal

Abstract

The importance of Computer Assisted Decision systems is of ever growing importance for gastroenterology imaging scenarios. Automatic classification of cancer lesions are not a trivial task and new methodologies are needed. In this paper we perform the classification of chromoendoscopy images into three different groups using Haralick texture features and Neural Networks. The algorithm can be extended to embody a Multiresolution classification scheme, and the results show that this method has the potential to overcome current state-of-the art classifiers in gastroenterology imaging context.

1 Introduction

The most important goal in Gastroentetology (GE) imaging is to detect cancer effectively and provide a diagnosis while the disease is curable and in a resettable stage. While manual diagnosis still presents some problems (such as the duration of the procedure, expensive human resources and in some cases different diagnosis even by expert gastroenterologists) it is believed that Computer Assisted Decision (CAD) systems can mitigate those shortcomings and potentially improve the diagnostic capabilities, by reducing the analysis time, helping in training of physicians and providing a second opinion. In this paper we have focused on classification of Chromoendoscopy (CH) images, which use the full visible spectrum and produce images having rich color characteristics, obtained by the application of dyes combined with magnified and high resolution endoscopy. CH, just like other gastroenterology image modalities, have texture as a common visual feature. Previous works [1, 2] expose the superiority of texture features for classification of GE images as compared to color features. Furthermore, Multiresolution (MR) methods have been shown to provide significant improvements to most of applications based on texture features. In particular, it was demonstrated that the addition of a generic classification system, dubbed as MR Block ([4, 7, 8]), which performs the extraction of features in MR subspaces produced a statistically significant jump in the classification accuracy of protein subcellular location images. In this work we focus on Haralick texture features and perform the classification of our dataset using an Artificial Neural Network (ANN). The main purpose of this work is to pave a way towards a Multiresolution Approach in a GE imaging context.

2 Methods

2.1 Feature Extraction

Texture features were found to be superior in comparison to color features, which motivates the set of features used in this work (we did not use wavelet or Gabor filters, since they are inherently multiresolution, which will be applied in the future to this algorithm). Haralick texture features [3] are computed from various statistical properties by the combination of four gray-level co-ocurrence matrices. Given a $N \times N$ image $G(i,j)$ we define $P(i,j)$ as the number of times a pixel with grey-level i occurs within a given distance from a pixel with grey-level j. The four matrices represent horizontal (H), vertical (V), left diagonal (LD) and right diagonal (RD) nearest neighbours. Haralick defines 13 measures and our first feature set is calculated as the mean of each measure, resulting in 13 features $(i = 1, ..., 13)$

$$F_1(i) = \frac{F_H(i) + F_V(i) + F_{LD}(i) + F_{RD}(i)}{4}. \quad (1)$$

The second Haralick feature set is defined as [4]

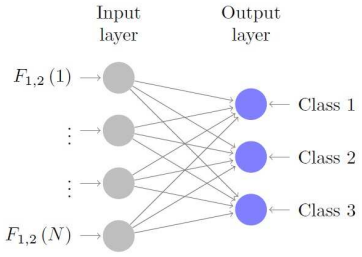


Figure 1: Neural Network topology [4] used in this work for classification. The input layer has the same number of nodes as the number N of extracted features (13 for F_1 and 26 for F_2). For the output layer each node correspond to a class (3 in total).

		Automatic Classification		
		Class 1	Class 2	Class 3
Manual annotations	Class 1	50	6	0
	Class 2	7	88	1
	Class 3	2	18	4

Table 1: Confusion matrix using the F_1 Haralick texture features (overall accuracy of 80,7%).

$$F_2(i) = \frac{F_H(i) + F_V(i)}{2} \quad F_2(i+13) = \frac{F_{LD}(i) + F_{RD}(i)}{2} \quad (2)$$

resulting in a total of 26 features.

2.2 Classification

After the extraction of texture features from the images, we proceed to the classification of our dataset, where we classify each image into three possible categories (normal, pre-cancer and cancer). In this work we constructed a Feedforward ANN [6] for the classification task. In general ANN's are composed by an input layer, one or more hidden layers and an output layer. The number of basis functions is fixed but are allowed to be adaptive on the training process by the minimization of an error function. The usual way to minimize it is through the Backpropagation method. We tested the classification accuracy for a topology consisting on one input layer with an hyperbolic tangent-sigmoid transfer function and one output layer with a linear transfer function (no hidden layers) following [4] dubbed as "Base System" (Figure 1.).

		Automatic Classification		
		Class 1	Class 2	Class 3
Manual annotations	Class 1	50	6	0
	Class 2	5	91	0
	Class 3	4	14	6

Table 2: Confusion matrix using the F_2 Haralick texture features (overall accuracy of 83,5%).

		Automatic Classification		
		Class 1	Class 2	Class 3
Manual annotations	Class 1	43	13	0
	Class 2	2	93	1
	Class 3	5	12	7

Table 3: Confusion matrix using the *LBP* (overall accuracy of 81,3%).

3 Material - Chromoendoscopy Images

The CH dataset consists on 176 images annotated by expert physicians, and each image belong to one of the classes: "Normal" (Class 1), "Pre-cancer" (Class 2) and "Cancerous" (Class 3). Those images were obtained using an Olympus GIFH180 endoscope at the Portuguese Institute of Oncology (IPO) Porto, Portugal during routine clinical work and the endoscopic videos were recorded during real endoscopic examinations.

4 Experimental Results

For our experiments the feature extraction stage was performed using the Haralick texture features [9] F_1 and F_2. The Regions of Interest (ROI) annotated by physicians are assumed. For the classification stage we have used a Neural Network with the topology described in Section 2.2 (see also Figure 1) [6]. The nodes in the second layer (output layer) have a linear transfer function and each one corresponds to a class (3 in total) an in order to avoid over-fitting we choose the ANN parameters correspondent to the smallest error with respect to the test data (early stopping). We also performed a 5-fold cross validation and the confusion matrices were added up in the end of the five interactions. The overall accuracy achieved is 80,7% for F_1 (Table 1.) and 83,5% for F_2 (Table 2.). Although our result using F_2 is higher than the accuracy obtained by [2], the images for that work were selected individually for the study by an expert clinician in order to avoid intra-patient repetition and badly captured images, resulting in less 12 images (164 in total). Nonetheless we show that by using Haralick texture features alone we get a similar result when compared with Local Binary Patterns (LBP) and sets of features based on Gabor filters (which are inherently of multiresolution nature). As we can observe from the confusion matrices, the system has the ability to correctly classify images belonging to Class 2 with true positives around 92% and 95%. This can be explained by the fact that most of the data set belongs to this class (54,5% of whole dataset) which unbalance the training ability of the ANN. Another ANN topology worked well for LBP (overall accuracy of 81,3%, Table 3.) and shows their general importance for the classification of CH images [10].

5 Discussion and Future Work

In this paper we established the first step for the application of a Multiresolution approach in a GE scenario by testing an ANN classification system. The performance of Haralick texture features on this dataset is comparable with the Gabor Energy feature extraction in [2], which is a multiresolution technique. It is expected an increase of the overall accuracy of this paper with a MR classification system considering the results of several authors ([4, 7, 8]). The implementation of this method and the construction of specific wavelet decompositions (and ANN topologies) suited for GE images is left for future work.

6 Acknowledgements

This work is funded by Instituto de Telecomunicações in the scope of Project BRIDGE, Ref[a] CMUP-EPB/TIC/0069/2013. Vasco Ferreira would also like to thank Fábio Domingues for the CPU availability.

References

[1] M. Coimbra et al., Segmentation for classification of gastroenterology images, in Proc. of IEEE EMBC, Buenos Aires, Argentina, Sep. 2010.

[2] F. Riaz et al., Gabor textons for classification of Gastroenterology Images, in IEEE International Symposium on Biomedical Imaging, 117 - 120, Chicago, U.S.A, 2011.

[3] R.M. Haralick, "Statistical and structural approaches to texture", in Proc. IEEE, 67:786-804, 1979.

[4] Amina Chebira et al., "A multiresolution approach to automated classification of protein subcellular location images", in BMC Bioinformatics 2007, 8:210.

[6] Tom Schaul et al., "PyBrain", in Journal of Machine Learning Research 2010, 11 743-746.

[7] Srinivasa G. et al., "Adaptive multiresolution techniques for subcellular protein localization image classification", in Proc. IEEE Int Conf Acoust, Speech and Signal Proc Volume V. Tolouse, France; 2006:1177-1180.

[8] Merryman T. et al., "A multiresolution enhancement to generic classifiers of subcellular protein localization images", in Proc. IEEE Int Symp. Biomed. Imaging Arlington, VA; 2006:570-573.

[9] Coelho, L.P. 2013. Mahotas: Open source software for scriptable computer vision. Journal of Open Research Software 1(1):e3, DOI: http://dx.doi.org/10.5334/jors.ac.

[10] A. Sousa et al., "Identifying cancer regions in vital-stained magnification endoscopy images using adapted color histograms", in Proc. of. IEEE ICIP, Cairo, Egypt, Nov 2009.

SABADO - SmArt BrAnd DetectiOn

Gonçalo Oliveira[1]
gpalaio@student.dei.uc.pt
Bernardete Ribeiro[1]
www.dei.uc.pt/~bribeiro
André Pimentel[2]
andre.pimentel@eyesee.pt
Xavier Frazão[2]
xavier.frazao@eyeseesolutions.com

[1] CISUC – Department of Informatics Engineering
University of Coimbra
Coimbra, PT
[2] EyeSee Solutions
Av. 5 de Outubro nº 293 4th Floor
Lisbon, PT

Abstract

We present an automatic graphic logo detection system that robustly handles unconstrained imaging conditions. The system is based on the novel Fast Region-based Convolutional Networks (FRCN) model, proposed by Ross Girshick, which has shown state-of-the-art performance in several generic object recognition challenges (PASCAL Visual Object Classes challenges). The novelty lies in the use of transfer learning to leverage powerful Convolutional Neural Network models trained with large-scale datasets and repurpose them in the context of graphic logo detection. Another benefit of this framework is that allows for multiple detections of graphic logos using regions that are likely to have an object. Experimental results with the FlickrLogos-32 dataset show not only the promising performance of our method with respect to noise and other transformations a graphic logo can be subject to, but also its superiority over state-of-the-art systems with hand-crafted models and features.

1 Introduction

In order to captivate their customers and make better decisions, Brands have the need to analyze the presence of their brand in images and other types of content. Brand logos help to assess the identity of something or someone. Most solutions use graphic logos as the main target of detection since they often present distinct shapes and appear in high contrast regions. It is a challenge since they are often subject to multiple angles and sizes, varying lighting conditions and noise.

Most previous works in this context have been based considerably on the use of SIFT [6]. This method provides representations and transformations to image gradients that are invariant to affine transformations and robust when facing lighting conditions and clutter. They can also detect stable salient points in the image across multiple scales, usually called key-points. These previous works build models upon these representations to better capture specific patterns present in graphic logos. For instance Romberg et al. [8] propose a shape representation built with found key-points and their respective SIFT representation. Similarly, Romberg and Lienhart [7] build bundles of SIFT features from local regions around each key-point to index specific graphic logo patterns.

Convolutional Neural Networks (CNNs) are multi-layer neural networks designed to recognize visual patterns directly from image pixels. The work by LeCun et al. [5] has been one of the main pioneering works for the current CNNs that are researched today. Recently CNNs have been in the center of object recognition research. The rekindled interest in CNNs is largely attributed to Krizhevsky et al. [4] CNN model, that showed significantly higher image classification accuracy on the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Their success resulted from a model inspired by LeCun previous work and a few twists that enabled training with 1.2 million labeled images (*e.g.* GPU programming, max(x,0) rectifying non-linearities and "dropout" regularization).

Our work is focused on providing a way towards graphic logo detection by utilizing general region proposal algorithms and state-of-the-art object recognition systems. However, due to the lack of a large scale dataset with such graphic logos, training a modern system with CNNs from the scratch is mostly prohibitive, therefore, we use transfer learning and data augmentation to ameliorate this problem.

2 Our approach and contributions

We extend and build upon general concepts and models of object recognition in images. Thus, we propose a solution that takes advantage of specific characteristics of graphic logos based on current cutting-edge research. The system we propose uses transfer learning to leverage image representations learned with CNNs on large-scale annotated datasets. The transferred representation leads to significantly improved results for brand detection. This is a very important innovation in this study because we have shown empirically that it performs well and can be as well applied in other contexts and problems.

3 Transfer learning with CNNs

The convolutional layer of a CNN consists of a set of learnable filters that activate with specific image features. Earlier layers of these networks learn to detect generic features like edges, and as we move into further layers, the learned features get more and more specific towards the original dataset. It is then possible to continue training and repurpose, or transfer the weights so that they adapt to other datasets. Several authors have shown that fine-tuning these networks and transferring features, even from tasks that are not closely similar, can be advantageous when compared to training from the scratch [1]. The process takes advantage of the more generic learned features, given that, both datasets live in a similar domain, in this case, object detection in images. This method enables training a large network with a small dataset without overfitting.

4 Fast Region-based Convolutional Networks (FRCN)

Graphic logos are not usually the main focus of an image, so they are often present in small sizes and partially occluded. Performing classification using the full image would introduce high amounts of background noise, given that fact, performing a more exhaustive search is required in this context. Some methods are now employing region proposals to reduce the initial search space and perform detection using powerful CNN models.

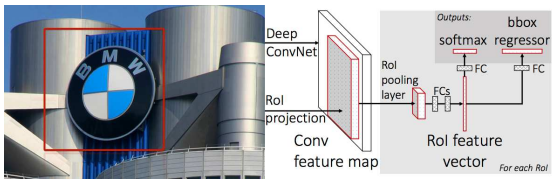


Figure 1: FRCN model. Adapted from [2].

Girshick [2] proposes a single-stage object detection model using a region proposal algorithm and CNN features. It performs object detection by classifying multiple regions in the image. This method has shown state-of-the-art performance in PASCAL VOC 2012. Its high level representation is shown in Figure 1. Using category-independent region proposals, the model performs classification in a reduced set of regions that are likely to have present an object, this helps to reduce the potential false-positives and the number of regions to analyze. In particular, it uses the approach by Uijlings et al. [9]. This method generates regions using clustering of similar regions and several diversification methods, capturing possible regions where an object might be without the use of

sliding windows. Feature extraction is performed using a CNN for each of the proposed regions and classified using a softmax classifier with a Fully Connected (FC) layer. Duplicate region detections are reduced with non-maximum suppression and localization of the object is further refined using bounding-box regressors.

5 Evaluation

We evaluate the FRCN model for brand detection with the FlickrLogos-32 dataset, which contains 32 graphic logo classes. The training set contains 320 images, the validation set 960 and the test set 960 images, each showing at least a single logo. Additionally, the test set contains 3000 images that do not contain any logo from the 32 classes. The recognition precision and recognition recall are the main scores used to measure performance on this dataset, although we will also evaluate the detection precision and recall individually. In this context, recognition consists of correctly classifying the class of images with graphic logos and detection consists of correctly classifying if a graphic logo is present or not. We perform data augmentation by horizontally flipping the training images and since the training set only contains 10 images per class we also use the original validation set for training. We use the Caffe deep learning framework [3] implementation of the FRCN model and produce results using two pre-trained CNNs. The first is the Caffe implementation of the CNN described by Krizhevsky et al. [4] called Caffenet, the second, a larger model, which has the same depth but wider convolutional layers, VGG_CNN_M_1024, described in [1]. Both were pre-trained with the ILSVRC dataset. The training process of the FRCN model jointly optimizes a softmax classifier and bounding-box regressors while adapting the CNN weights to our new task (graphic logo detection). We start Stochastic Gradient Descent at a small learning rate of 0.001, which allows fine-tuning of the CNN network to occur without losing the capabilities gained during pre-training. We fine-tune Caffenet during 30000 iterations and VGG_CNN_M_1024 during 40000. We alternatively designate the Caffenet model as **M1**, and the larger model, VGG_CNN_M_1024 as **M2**. We will abbreviate the name of our final brand detection model to **SBD-FRCN**.

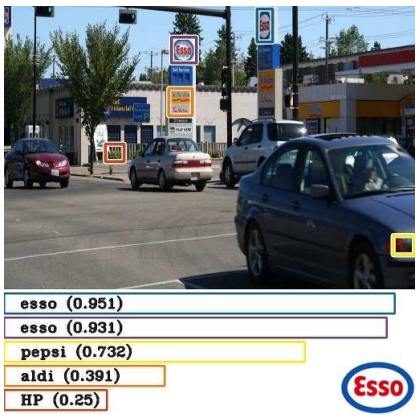


Figure 2: Top-5 confidence scores for an example image belonging to the Esso class.

The softmax classifier will give a confidence value for each class, throughout all proposed regions, we will use the region with top confidence value to classify each image. Figure 2 shows the top-5 confidence scores produced by this system for a test image.

We introduce a threshold value to deal with images without graphic logos (no-logo). Images with top confidence values below the threshold will be considered as having no logo present. Figure 3 shows consolidated F1-scores for both detection and recognition metrics and the two tested CNN models. Using a threshold value of 0.3 we achieve the top recognition F1-score, although by setting an adequate threshold value still allows the model to maintain high recognition scores without losing the ability to perform detection between images with logos and no graphic logos (*e.g* threshold of 0.8). Table 1 also shows the best results achieved by other authors on this specific dataset that we know of, as we can see, our results with this model compare favorably.

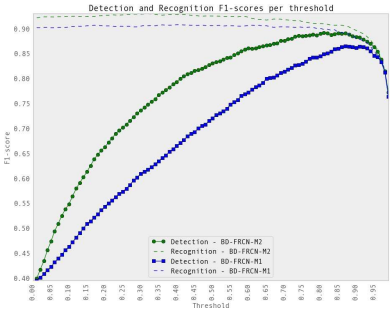


Figure 3: F1-scores per threshold value, for both SBD-FRCN-M2 and BD-FRCN-M1.

Method	Precision	Recall	F1
Romberg et al. [8]	0.982	0.61	0.753
Romberg and Lienhart [7]	0.999	0.832	0.908
SBD-FRCN-M1 (thresh 0.4)	**0.928**	**0.891**	**0.909**
SBD-FRCN-M2 (thresh 0.81)	**0.987**	**0.846**	**0.911**
SBD-FRCN-M2 (thresh 0.32)	**0.955**	**0.908**	**0.931**

Table 1: Recognition scores.

6 Conclusion

A key contribution of this work has been the introduction of graphic logo detection using regions and CNNs, taking advantage of transfer learning. We experimented with a modern detection model and two CNNs pre-trained for a general object recognition task with abundant data and fine-tuned these networks for a task where the data is scarce (graphic logo detection). The model we propose compares favorably to state-of-the-art performance almost out of the box, with a wide margin for improvement that we intend to explore in the future.

7 Acknowledgements

We would like to thank Eng. João Redol for the initial feedback as well as Eyesee Solutions for their support.

References

[1] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. *arXiv preprint arXiv:1405.3531*, 2014.

[2] Ross Girshick. Fast r-cnn. *arXiv preprint arXiv:1504.08083*, 2015.

[3] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proc. of the ACM Int. Conference on Multimedia*, pages 675–678. ACM, 2014.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

[6] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, Nov 2004.

[7] Stefan Romberg and Rainer Lienhart. Bundle min-hashing. *Int. Journal of Multimedia Information Retrieval*, 2(4):243–259, 2013.

[8] Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, and Roelof van Zwol. Scalable logo recognition in real-world images. In *Proc. of the 1st ACM Int. Conference on Multimedia Retrieval*, ICMR '11, pages 25:1–25:8, New York, NY, USA, 2011. ACM.

[9] Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *Int. journal of computer vision*, 104(2):154–171, 2013.

Glucose Level Prediction in Diabetic Patients

Daniel Frutuoso[1]
danielf@student.dei.uc.pt
Bernardete Ribeiro[1]
bribeiro@dei.uc.pt
André Pimentel[2]
andre.pimentel@eyesee.pt

[1] CISUC - Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal

[2] EyeSee
Av. 5 de Outubro Nº293, 4th floor
Lisbon, Portugal

Abstract

Diabetes is a huge health problem that is affecting more and more people over the time. Since there is still no cure for this disease, a person who has been diagnosed with diabetes has to control his glucose level. The power of machine learning techniques has been increasing allowing to solve more and more complex problems in medicine. This work aims to investigate how accurately Linear Regression (LR) and Support Vector Regression (SVR) models can be when predicting the glucose levels in diabetic patients by assessing the Root Mean Squared Error (RMSE). The obtained results show that LR performs better than SVR with an average RMSE of 13.56 mg/dL and 22.20 mg/dL when predicting 30 minutes and 60 minutes ahead, respectively.

1 Introduction

Diabetes is a major growing health problem affecting more and more people in the planet which may lead to death. It is a metabolic and chronic disease characterized by the absence or low production of insulin associated or not with a deficient action of it in the organism. There are symptoms associated with the unbalanced glucose levels. A person is on a hypoglycaemia state when the glucose level is below 70 mg/dl and if it is ignored it can lead to unconsciousness, permanent brain damage or death. If it is above 200 mg/dl two hours after eating or greater than 126 mg/dL when fasting, the subject is considered to be in a hyperglycaemia state, which can conduct to cardiovascular disease, blindness, kidney failure or damage, nerve damage, etc.

World Health Organization (WHO) classified diabetes as one of the 10 top causes of death between the year 2000 and 2012 and in 2030 it is expected that it will be the seventh[1]. Prediction of future glucose concentrations is a crucial task for diabetes management which may help diabetics from entering on hypoglycaemia and/or hyperglycaemia events and suffering its consequences.

2 Computational Experiments

In this section, a description of the datasets and how they were handled is given. It also presents two prediction methods and the RMSE evaluation metric to assess the prediction models.

2.1 Datasets

In order to create computational models capable of predicting diabetics' glucose levels, we used 5 real datasets: CGM, RA, AML1, AC1 and RR. The first one can be found in Zaitlen et al. [4] while the other four were given by Associação Protectora dos Diabéticos de Portugal (APDP) through a collaboration protocol. Each one of them keeps records of patients' glucose concentration in a quasi-continuous way, during some period of time, which allows a better insight of the variation of the glucose level. Those files were obtained by the use of a continuous glucose monitoring device with sampling intervals of 5 minutes that is placed on the subjects belly. In the CGM file, there are missing values which lead us to choose the biggest period without gaps. The remaining ones have the glucose level of each patient for almost 7 days.

2.2 Data pre-processing

For every dataset, data was normalized using the min-max normalization technique which maps a value x to x' in the range $[0,1]$. Since glucose level may oscillate between 40 and 600, it is possible to use this technique. However, to make sure that no new value would appear, that range was enlarged by setting the minimum possible value to 20 and the maximum to 800.

After normalizing data and to use it to build the computational models, there is the need to organize data in the proper format. Taking into account that we only have information about the glucose level, the model will predict future values based on past glucose measurements. Considering $z = f(t)$, which represents the glucose level at instance t, the value is mapped to the glucose level at target time $y = f(t+PH)$, where PH is the prediction horizon, making our dataset $D = \{(z_1, y_1), (z_2, y_2), ..., (z_N, y_N)\}$. Although it is possible to train a model using pairs of $(\mathbf{z}_i, y_i)$, it is expected that if we include more information, the model will yield better results. In other words, instead of relying just on the present glucose level one might enlarge the number of past values. Given that we extended each pattern with W past measurements, $\mathbf{z}$ is now $\{f(t), f(t-1), f(t-2), ..., f(t-W)\}$. For this reason, we transformed each of the original dataset into another datasets taking into account the different values of W and PH.

Finally, data is divided into a training set and a testing set. The first one corresponds to 70% of all data while the second one corresponds to the remaining 30%.

2.3 Prediction Methods

To predict the glucose levels on a individual there are some multi-step-ahead prediction techniques, that given a time series $\{z_1, z_2, ..., z_N\}$ composed of N observations consist on predicting $\{z_{N+1}, z_{N+2}, ..., z_{N+PH}\}$, where $PH \geq 1$.

In this section, the two prediction strategies [3, 5] used in this work are described: iterative prediction method and direct prediction method.

2.3.1 Iterative Prediction method

In this method, once a one-step-ahead prediction is computed, the value is given again as an input to the next step following a recursive strategy.

The one-step ahead prediction model depending on W past points has the form

$$\widehat{z}_{t+1} = f(z_t, z_{t-1}, ..., z_{t-W+1}) \tag{1}$$

After the value of $\widehat{z}_{t+1}$ has been computed it is used to predict $\widehat{z}_{t+2}$ and so on until the desired PH. In general, the estimation of the PH next values is returned by

$$\begin{cases} \widehat{z}_{t+1} = f(z_t, z_{t-1}, ..., z_{t-W+1}) \\ \widehat{z}_{t+2} = f(\widehat{z}_{t+1}, z_t, ..., z_{t-W+2}) \\ ... \\ \widehat{z}_{t+PH} = f(\widehat{z}_{t+PH-1}, \widehat{z}_{t+PH-2}, ..., \widehat{z}_{t+PH-W}) \end{cases} \tag{2}$$

From Equations 2 it is possible to conclude that to predict the value for $\widehat{z}_{t+PH}$ it requires PH one-step ahead predictions.

2.3.2 Direct Prediction method

An alternative to the iterative method is the direct prediction method which for PH-steps ahead the value is given by

$$\widehat{z}_{t+PH} = f_{ph}(z_t, z_{t-1}, ..., z_{t-W+1}) \tag{3}$$

where $1 \leq ph \leq PH$ and f_{ph} is a model that predicts directly the value at instant $t+ph$ without any predicted values. In general, the prediction of the PH next values is returned by

[1] http://www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf

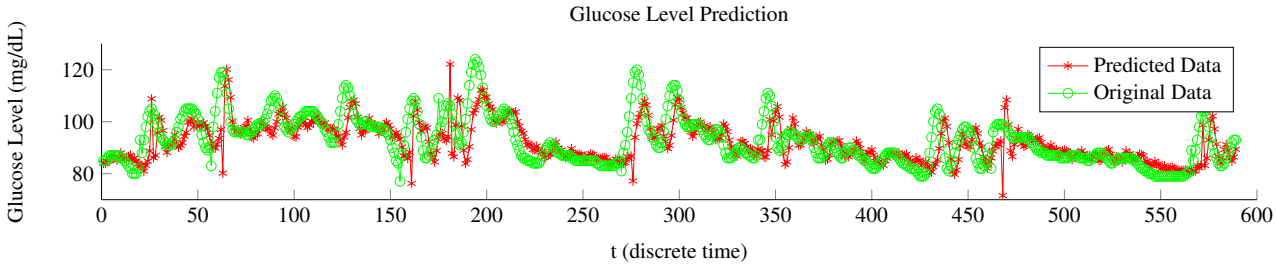


Figure 1: Prediction of glucose levels 30 minutes ahead using Linear Regression with direct prediction method - RA file

Algorithm	Kernel	Prediction Method	RMSE I	RMSE II
LR	–	Direct	13.56	22.20
	–	Iterative	13.50	22.00
SVR	Linear	Direct	13.99	22.78
		Iterative	13.85	22.82
	RBF	Direct	13.70	22.67
		Iterative	13.87	22.94

Table 1: Average RMSE of the 5 datasets in mg/dL when predicting (I) 30 minutes ahead and (II) 60 minutes ahead

$$\begin{cases} \widehat{z}_{t+1} = f_1(z_t, z_{t-1}, ..., z_{t-W+1}) \\ \widehat{z}_{t+2} = f_2(z_t, z_{t-1}, ..., z_{t-W+1}) \\ ... \\ \widehat{z}_{t+PH} = f_{PH}(z_t, z_{t-1}, ..., z_{t-W+1}) \end{cases} \quad (4)$$

To predict all the values from $\widehat{z}_{t+1}$ to $\widehat{z}_{t+PH}$, PH models need to be built. However, it is possible to use just one of those models accordingly to the chosen PH, if the model gets new data with the same interval as it needs to predict. For example, considering that we are on instant t and $PH = 2$ with a sample time of 5 minutes, we would compute $\widehat{z}_{t+2}$ with $f_2(z_t, z_{t-1}, ..., z_{t-W+1})$. After 5 minutes that corresponds to instant $t+1$ the value of z_{t+1} on instant t is available which now becomes the value of z_t and can be used to predict $\widehat{z}_{t+3}$ with f_2.

2.4 Evaluation Metric

In order to check and compare the performance of the generated models, we have to calculate a metric to do so.

Root Mean Squared Error (RMSE) represents the standard deviation of the differences between the real values and the observed ones. This metric has the advantage that it is an easily interpretable statistic since it has the same units as the data plotted on the vertical axis. RMSE is a good measure of how accurately the model predicts the response and the smaller the value is, the better the model.

3 Results and Discussion

In this section, the results of the experiments are presented and discussed. Table 1 shows the average results when predicting the glucose level of the 5 patients 30 and 60 min ahead with both prediction methods.

The first conclusion that it is possible to take is that the model built using the linear regression algorithm performed better when compared with SVR in both prediction horizons. There is a linear relationship between the previous glucose levels and the output since both linear regression and SVR with linear kernel provide good results. Supposedly, SVR with RBF kernel ought to perform better than the one with the linear kernel given yet it does not always happen, which reinforces the theory of a linear relationship.

When it comes to the prediction methods, neither one is significantly better being the maximum RMSE difference of just 0.27 mg/dL, which has a residual effect on the final predicted value and does not affect the patient. For this reason and following the KISS principle[2] , we chose the direct prediction method as with it, less computational power is needed.

[2]KISS - Keep It Simple ...

In Figure 1, the true level of glucose of the RA patient is compared against the predicted values given by the linear regression model with the direct prediction method. A perfect model would have both lines overlapped. However, from the graphic we conclude that the model is capable of following the dynamics of the glucose.

Comparing the results of the other works, our results outperformed those. Georga et al. [1] used the SVR algorithm achieving a RMSE average of 16.92 mg/dL and 25.02 mg/dL predicting 30 and 60 minutes ahead. Marling et al. [2] also tried to predict the glucose levels with the SVR algorithm accomplishing a RMSE of 18.00 mg/dL and 30.9 mg/dL for $PH = 30$ and $PH = 60$ min, respectively. From the clinical point of view, a model that has the lowest RMSE is the best since its predictions would be closest to the real future values and would not give false alerts to the patient. Given that, our model is superior even for $PH = 60$ minutes.

4 Conclusion

Diabetes is a serious health problem that affects millions of individuals and has several consequences if it is not controlled. In this work, we compared the performance of two algorithms predicting the diabetics' glucose level. The results showed us that the best model to predict future glucose levels was the one built using the linear regression algorithm. Although it is a simple one when compared to others like the SVR, it was possible to see that it achieved better RMSE than the other one when predicting 30 and 60 minutes ahead. In terms of the prediction method, it is not possible to say that one is better than the other as there is no significant difference in the results. However, since the direct method requires fewer calculations we consider it better in a real case scenario. In this work, we managed to predict the subjects' glucose level with an average error of 13.56 mg/dL and 22.20 mg/dL, respectively, which may prevent them from entering on hypoglycaemia or hyperglycaemia states.

Acknowledgements

The authors would like to thank EyeSee Solutions for the support during this project.

References

[1] E. Georga, V. Protopappas, D. Polyzos, and D. Fotiadis. Predictive modeling of glucose metabolism using free-living data of type 1 diabetic patients. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual Int. Conf. of the IEEE*, pages 589 – 592, Aug 2010.

[2] C. Marling, M. Wiley, R. Bunescu, J. Shubrook, and F. Schwartz. Emerging applications for intelligent diabetes management. *AI Magazine*, 33(2):67, 2012.

[3] S. Taieb, A. Sorjamaa, and G. Bontempi. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10 - 12):1950 – 1957, 2010.

[4] B. Zaitlen, L. Desborough, and A. Ahmadia. Continuos Glucose Monitor Data. 07 2013. URL `http://dx.doi.org/10.6084/m9.figshare.741296`.

[5] L. Zhang, W. Zhou, P. Chang, J. Yang, and F. Li. Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, 99(0):411 – 422, 2013.

Analysis of Pedestrian Models for Social Behaviour Discovery in Surveillance Scenario

Eduardo M. Pereira
ejmp@inescporto.pt
Jaime S. Cardoso
jaime.cardoso@inescporto.pt

INESC TEC
Campus da FEUP, Rua Dr. Roberto Frias,
4200 - 465 Porto, Portugal

Abstract

Modelling human activity within a sociological principled way has an undeniable value for both low-level problems such as pedestrian tracking, and high-level applications such as anomaly detection in security. This work brings a new perspective to investigate the role of pedestrian models for social behaviour analysis. Several conclusions in term of accuracy to simulate real social behaviour, alternatives to evaluate the annotation process and pertinence of simulated data on social analysis are taken.

1 Introduction

Automatic behaviour understanding from CCTV (Closed-circuit television) systems is a very complicated problem. It comprises several hierarchical layers of processing, from bottom low-level features to top high-level semantics interpretation. Multi-target tracking is one of those low-level critical steps, and normally relies on a dynamic model to predict the object's location from its previous past information. Most common methods predict the new location of each target based on its history and focus their contribution on improving the appearance object model, and corresponding detection, or the optimisation strategy to find correct object's assignments [2]. The knowledge of high-level scene surrounding such as collisions, proxemic distances and social interactions, has been introduced in recent works to improve tracking performance, specially during occlusions [6]. In the literature, collective behaviour analysis tend to fit into two types of taxonomy: the one that considers groups as a collective and homogeneous block where individual is transformed by the group, the so-called *macroscopic* studies [10], and the one that analyses groups as the composition of individual agents that interact with each other and with the environment, the *microscopic* approaches [4]. The latter simulates pedestrian physical behaviour using cues such as relational connections among people, focus of attention of each person, geometric scene constraints, and proxemics-based distances.

For our specific scenario, *microscopic* studies are more suitable but their formulation is not enough to derive social semantic behaviour. Therefore, we formulated a social behaviour framework that represent and classify individual profiles (I.P.s) and group behaviours (G.B.s). In this work, we simulate different pedestrian models and embedded their results into our framework to inspect their accuracy to imitate real-life social behaviours. The evaluation is three-fold: i) selection and discussion of the pedestrian model that best describe the manual annotation; ii) inspection of the robustness and coherence of the annotation process; iii) adaptability of our framework to identify I.P.s under different social parameters simulated. A brief study about the impact of the number of agents on the social behaviour classification is elaborated. As well, the advantages of the performed analysis to improve tracking, increase social knowledge representation, and automate the annotation process and alleviate its ambiguity are highlighted.

2 Background

2.1 Semantic Concepts and Dataset

We selected the IIT (Israel Institute of Technology) dataset and were granted access by the authors [1]. The chosen scenario is a shopping mall (83155 frames with resolution 512×384 @25 fps). We were advised by the lab of social-psychology of the University of Porto to identify and annotate the I.P.s and G.B.s. The dataset, including annotation, is available upon request. The semantics in terms of I.P.s and G.B.s were defined considering the environment as social context. The I.P. follows a grammar-based analogy where an individual action is associated to an adjective. The G.B. adopts a definition of group dynamics that explains the interdependence degree among individuals and their influence over the group behaviour they belong to. In this work we just take into account the I.P.s, which were specified as: i) *exploring (Exp.)*, when no specific interest is revealed, but movement and gaze are coherent with the scene structure and context; ii) *interested (Int.)*, when an interest by an object on the scene is explicitly revealed; iii) *distracted (Dist.)*, when no specific interest is revealed, translated into unstructured movement and variability on gaze; iv) *disoriented (Dis.)*, when confusion concerning interests is revealed, expressed into a high variability on movement and gaze along with an structured movement. Please, refer to [7] for a more detailed overview of the dataset characteristics, as well as the explanation of the novel semantic concepts and annotation considerations.

2.2 Social Behaviour Framework

The social behaviour framework encompasses the identification of position and attention-based features, the formulation of a relational social descriptor and a Bag-of-Features (BoF) classification approach with different sampling, pooling and feature matching techniques.

The descriptor is composed by several features, which are encoded into a multi-scale histogram controlled by R, the number of granularity levels where the number of bins are given by 2^R, and the final descriptor is the concatenation of each feature's histogram. The descriptor is fixed-length to be embedded into a BoF classification approach. The codebook was build by running k-means over a subset of the annotated data. The centers form the vocabulary to be used on further training and classification processes. The sampling follows a key-point trajectory strategy, where each descriptor is extracted over a temporal gap, τ, expressed in seconds. Each bag is composed by consecutive descriptors and its length is controlled by Γ. A multi-class SVM classifier was trained to identify the different I.P.s and G.B.s In terms of I.P.s the following features were taken: i) *angular direction change*, α_{si}, is the angular variation of movement of the individual between consecutive sampling times. ii) *distance of interest*, d_{io}, expresses the distance between individual position and the object of interest. iii) *direction of interest*, β_{gi}, which is the gaze direction. iv) *velocity*, v_i, expresses the instantaneous velocity of the individual at the sampling time. Please, refer to [7] for more information.

3 Pedestrian Models and Simulation

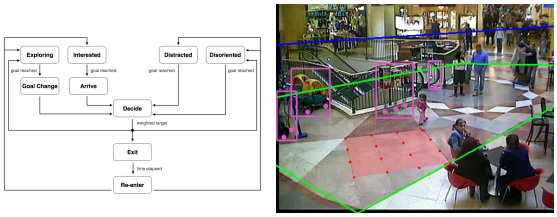
To model the behaviour of pedestrians, we used the Menge [1] crowd framework [3]. Considering the knowledge of the environment such as obstacles and surrounding neighbours, the navigation plan is adapted locally to transform the desired velocity into a feasible velocity. This local navigation is controlled by the pedestrian models that follow a *microscopic* approach. In this work, we simulated the motion dynamics adaptation phase with different pedestrian models: i) *pedvo*, based on pedestrian velocity obstacles; ii) *orca* [8], a model that considers local collision avoidance; iii) *helbing* [4], a simple social-force model; iv) *johansson* [5], an evolutionary social-force model that adjust its parameters accordingly with tracking history; v) *zanlungo* [9], another social-force model that explicitly handle collision prediction.

The agent behaviour and how it changes are described by a Behavioural Finite State Machine (BFSM). Considering the I.P.s semantics stated in Section 2.1, our simulation follows the chain of states illustrated in Figure 1(a). The real mall scenario (see Figure 1(b)) was recreated taking in consideration the ground plane dimensions, the objects of interest, and the entry/exit areas, and a navigation mesh was designed. The states that correspond to our I.P.s were modelled accordingly with two parameters, speed and gaze, which affect the position and attention-based features mentioned in Section 2.2. Several statistics, such as *min*, *max*, *mean* and *stddev*, were taken from the manual annotation trajectories and gazes for each I.P., and gaussian and uniform distributions were used to cope speed

[1] http://gamma.cs.unc.edu/Menge/

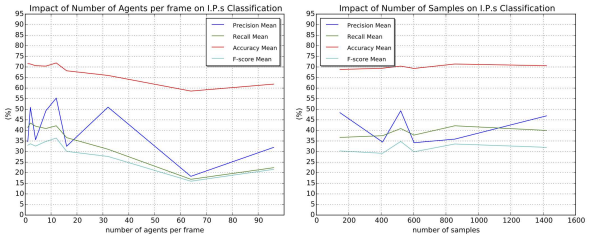
	Dist.				Exp.				Dis.				Int.				Avg.			
	P	R	A	F	P	R	A	F	P	R	A	F	P	R	A	F	P	R	A	F
manual	**32.0**	**51.1**	**89.8**	**39.0**	**97.2**	88.3	**87.2**	**92.5**	17.9	20.0	**98.8**	**36.6**	35.5	**68.2**	**93.9**	45.4	45.7	**56.9**	**92.4**	**53.3**
pedvo	15.1	7.7	66.1	10.2	33.4	97.8	52.9	49.8	50.0	6.5	75.9	**11.5**	**98.8**	51.6	*86.7*	**67.8**	**49.3**	*40.9*	*70.4*	*34.8*
orca	*25.7*	6.6	70.0	10.5	28.0	**98.3**	41.4	43.5	0.0	76.1	0.0	0.0	87.5	41.2	82.9	56.0	35.3	36.5	67.6	27.5
helbing	16.4	*10.6*	62.2	*12.9*	33.5	96.6	52.4	49.7	35.3	6.7	73.7	11.3	96.4	31.0	82.9	47.0	45.4	36.2	67.8	30.2
johansson	7.5	4.6	*66.6*	5.7	*36.8*	90.4	*53.7*	*52.3*	0.0	**76.7**	0.0	0.0	80.8	*53.2*	84.1	64.1	31.3	37.0	70.3	30.5
zanlungo	11.9	8.5	61.4	9.9	32.9	91.9	49.2	48.5	**66.7**	1.9	*77.8*	3.7	78.6	26.4	78.6	39.5	47.5	32.2	66.7	25.4

Table 1: Classification results for different pedestrian models using for training the annotation data and for testing the simulated data. The *manual* entry refers to the classification of the manual annotation accordingly with a 2-fold cross validation scheme with 10 random iterations.



(a) (b)

Figure 1: (a) BFSM used in our simulation; (b) Horizontal vanishing line (blue), ground plane's projection area (green), ground points (red), and objects of interest (purple).



(a) (b)

Figure 2: (a) Impact of the number of agents per frame in classification results; (b) Impact of the number of samples in classification results.

and gaze variations per state. We should mention that we extended the Menge framework to support gaze information. For that we used a simple weighted approximation that consider the previous gaze, a perturbation factor, and the current movement direction. Two type of goals were marked, the ones associated with the objects of interest, and some random goals defined along the free-walking area. The former are used by the *Int.* and *Exp.* profiles to mimic their attention to the scene's objects, while the latter are used by *Dist.*, *Dis.* and also *Exp.* to simulate random behaviour. We highlight that to the *Int.* state is attached an *Arrive* state, which consider a time interval to inspect the current object of interest. In the same way, the *Exp.* state is linked with a *Goal Change* state to afford the opportunity to change from scene's objects to random positions and vice-versa. The distance from the goals and the allowed time to spend around each one are also factors that our simulation took in consideration.

4 Experimental Results

Several experiments were performed to inspect: pedestrian models accuracy in terms of their availability to simulate social behaviours, ambiguity and coherence of the annotation process, and adaptability of our descriptor and classification framework to simulated data. To measure the accuracy to simulate social behaviours, we ran each pedestrian model with 8 agents per frame and collect their trajectories and gazes. That information was enclosed into our descriptor and the classification process was conducted considering as training set the manual annotation. Table 1 shows the results, where the first row refers to the classification of the manual annotation accordingly with a 2-fold cross validation scheme with 10 iterations. The *pedvo* model presents the best overall result, while the *orca* and *zanlungo* are the ones that perform worse. All of them incorporate collision avoidance mechanisms, the difference is that the *pedvo* model uses a velocity-based technique, which probably simulates better the social behaviour, while the others should approximate more the pedestrian to steering behaviours. As expected, the *helbing* and *johansson* present similar results, since both of them are based on social forces and their formulation just differ in the adjustment of parameters. Another important conclusion is that the *helbing* just present worse general results than the *pedvo* in the *Int.* profile, proving that it can also be a good choice.

Assuming the *pedvo* model, we investigate the impact factor on the classification results considering the variation of the number of agents per frame (Figure 2(a)), and the variation of the amount of samples to classify (Figure 2(b)), while training with the manual data. Our real scenario has an average of two pedestrian per frame, with a maximum of nine. We stated that the *pedvo* model presents it best results near those values. In terms of sampling, we verified a step variation around the 500 samples, and then a steady behaviour with a slightly increase on the precision.

Inverting the training and testing sets, Table 2, we verified that our annotation process is most similar with the *orca* model. We should correlate our annotation rules with the definition and rules of this model to take

	P	R	A	F
pedvo	29.6	39.2	62.3	17.6
orca	**29.9**	**41.6**	66.8	**41.6**
helbing	28.8	31.1	69.5	20.9
johansson	26.7	27.8	71.4	20.7
zanlungo	28.8	35.1	**72.3**	22.9

Table 2: Average classification results using for training the simulated data and for testing the manual annotation.

	P	R	A	F
pedvo	66.8	67.1	83.9	66.4
orca	**68.3**	**69.2**	**84.7**	**68.1**
helbing	67.9	68.0	83.7	67.3
johansson	64.6	64.7	83.4	63.9
zanlungo	63.8	63.8	82.6	63.1

Table 3: Average classification results of simulated data using a 2-fold cross validation scheme with 10 random iterations.

further conclusions. Table 3 shows, as expected, that the average classification results for each pedestrian model are very similar, since the data is simulated and follow well-defined rules. It also shows that our descriptor performs well under the characterisation of synthetic data.

5 Conclusions

This work presents preliminary results about a novel way to inspect the importance of pedestrian models in high-level inferring like social behaviour classification and its annotation process. We identified the most promising pedestrian model, took conclusions about the scalability of its performance, and verified the adaptability of our descriptor and classification framework for simulated data. However, we stated the difficulty in the simulation to tune different model parameters. We should also improve the update of the gaze parameter on each state in the BFSM.

Acknowledgements

This work is financed by the FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) within project UID/EEA/50014/2013, and for the PhD grant with reference SFRH/BD/51430/2011.

References

[1] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and David Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Trans. Pattern Anal. Mach. Intell.*, (3):555–560.

[2] Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. People-tracking-by-detection and people-detection-by-tracking. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2008. URL http://www.mis.tu-darmstadt.de/node/382.

[3] Sean Curtis, Andrew Best, and Dinesh Manocha. Menge: A modular framework for simulating crowd movement. *University of North Carolina at Chapel Hill, Tech. Rep*, 2014.

[4] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, (5):4282–4286, May . doi: 10.1103/physreve.51.4282.

[5] Anders Johansson, Dirk Helbing, and Pradyumn K. Shukla. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. *Advances in Complex Systems*, 10(supp02):271–288, 2007. doi: 10.1142/S0219525907001355. URL http://dx.doi.org/10.1142/S0219525907001355.

[6] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc J. Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *ICCV*, pages 261–268. IEEE.

[7] Eduardo M. Pereira, Lucian Ciobanu, and Jaime S. Cardoso. Context-based trajectory descriptor for human activity profiling. In *IProc. IEEE Int. Conf. Syst., Man, Cybern., San Diego, CA, USA*, October 2014.

[8] Jur van den Berg, Stephen J. Guy, Ming C. Lin, and Dinesh Manocha. Reciprocal n-body collision avoidance. In Cédric Pradalier, Roland Siegwart, and Gerhard Hirzinger, editors, *ISRR*, volume 70 of *Springer Tracts in Advanced Robotics*, pages 3–19. Springer, 2009. ISBN 978-3-642-19456-6. URL http://dblp.uni-trier.de/db/conf/isrr/isrr2009.html#BergGLM09.

[9] Zanlungo, F., Ikeda, T., and Kanda, T. Social force model with explicit collision prediction. *EPL*, 93(6): 68005, 2011. doi: 10.1209/0295-5075/93/68005. URL http://dx.doi.org/10.1209/0295-5075/93/68005.

[10] Bolei Zhou, Xiaogang Wang, and Xiaoou Tang. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *CVPR*, pages 2871–2878, 2012.

Dark slope streaks on Mars: Automated identification and temporal fading quantification

Fernanda Puga, Erivaldo A.Silva
ferpuga@gmail.com, erivaldo@fct.unesp.br

Faculdade de Ciências e Tecnologia,
Universidade Estadual Paulista, BRAZIL

Pedro Pina
ppina@tecnico.ulisboa.pt

CERENA, Instituto Superior Técnico,
Universidade de Lisboa, PORTUGAL

Abstract

We present an approach to identify and characterize temporally dark slope streaks on Mars: it starts by their segmentation, and then on the quantification of their temporal fading. The algorithms are successfully tested on remotely sensed imagery captured by different sensors (HiRISE, CTX and MOC) of different spatial resolutions (between 0.25 and 6.50 m/pixel) and in diverse geographical locations of planet Mars.

1 Introduction

Slope streaks are albedo features that occur often on current Mars [1]. They are typically dark, narrow and fan-shaped features that extend down slope and vary from a few to several hundreds of meters in length. Examples showing their diversity on Mars are provided in Figure 1.

Dry and wet processes have been suggested for causing their formation but their origin is still unclear [2]. For instance, some hypotheses indicate that they could have been formed by dust avalanches [3] or by melting frost or ice [4] as observed in analogue features in Antarctica [5].

Moreover, the streaks tend to fade becoming lighter with time and providing clues about dust deposition and target material properties. In order to understand the physical processes triggering these patterns, many surveys on Mars are being developed with remotely sensed imagery with the best resolutions available, namely, provided by the cameras HiRISE (up to 0.25 cm/pixel), MOC-NA (1.4 m/pixel), CTX (6 m/pixel) and also HRSC (12.5 m/pixel).

All the studies that quantify the characteristics of the streaks are based on manual delineations and by point sampling for computing the absolute albedo between them and the neighbouring regions. Consequently, the statistics are based on small amounts of data.

The only available methodology to segment the slope streaks and to extract meaningful information about them were recently initiated and proposed by ourselves [6-7], inspired by the segmentation of dust devils tracks also on Mars [8-9]. In this text we present the latest developments of our approach in the detection of these albedo features and the first steps for quantifying the temporal fading.

2 Identification algorithm

The slope streaks identification algorithm is based on its two main distinctive characteristics: 1) the albedo, since the large majority of the streaks are darker than the surrounding background and that the newly formed streaks are always darker than the older ones, and 2) the shape, as slope streaks are always narrow, fan-shaped and strongly elongated features. Thus, the algorithm is built to detect dark and elongated features and is constituted by three main steps: pre-processing, detection and post-processing. Each of these steps include the following operations:

i) Pre-processing - Consists in attenuating small bright/dark objects by a filtering operation based on morphological area-opening and area-closing operators [10]. The area-opening attenuates the reflectance of bright faces of structures like sand dunes and boulders that might be present on the scene. On the contrary, the area closing attenuates dark patches caused by shadows of small features and of the hills/slopes where the streaks were formed.

ii) Detection - The dark streaks are first enhanced through a top-hat transform by closing (also named as top-hat for valleys). The width (or thickness) is the parameter used by the top-hat transform [10] to distinguish dimensionally the slope streaks (the length is irrelevant for this transform). The dark regions or structures that do not completely contain the structuring element (we used a disk) are removed. Next, the thresholding with Otsu method [11] permits binarizing the images in an automatic way. After the segmentation it is still necessary to remove some small binary structures. Instead of using a classic erosion-reconstruction filtering, we use a thinning/pruning-reconstruction procedure [10]. This way we are able to keep the thinner structures, which would be suppressed by the erosion.

iii) Post-processing – It is based on the elongated shape of the streaks. The segmented objects that present a low value ratio 'length/width' are filtered out. The selection of an adequate threshold value is quite simple due to the high disproportion between those two measures. Thus, only a structure that has a length 5 times bigger that its width is preserved.

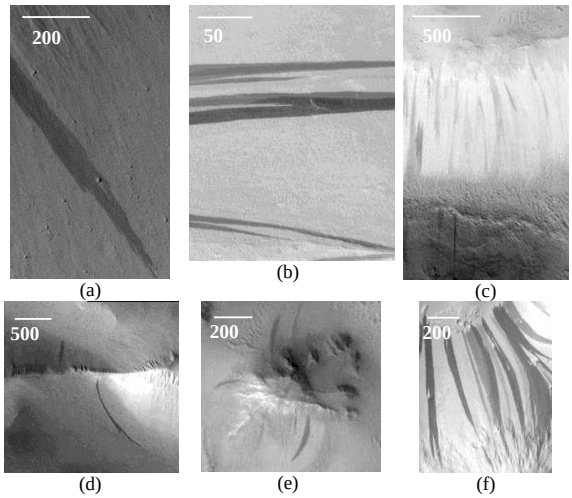


Figure 1: Dark slope streaks on Mars from images (a) HiRISE PSP_008513_2060, (b) HiRISE ESP_011730_2105, (c) MOC E1400638, (d) MOC M0307769, (e) MOC E1700689 and (f) MOC M0002117. The scale bars are meters.

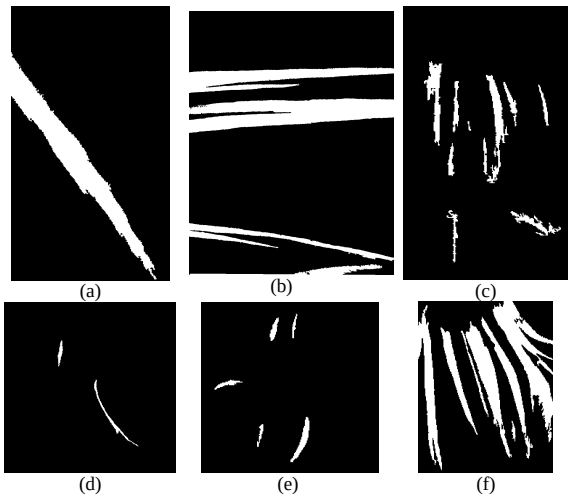
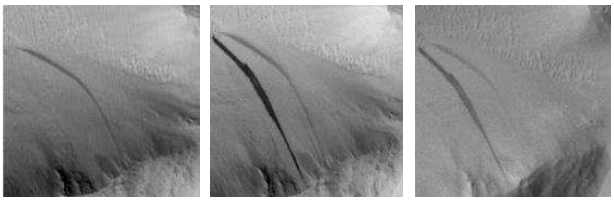


Figure 2: Final result of slope streaks detection of images in Figure 1.

3 Temporal fading quantification

This quantification requires a very good register between the images from different dates. In addition, the use of images captured by different cameras (also with distinct spatial resolutions) requires extra and careful efforts to get a good overlap. Our approach intends to not only quantify the temporal variation of the albedo of each individual streak until it completely fades out or disappears, but also to alert when new streaks are formed (see sequence in Figure 3).



(a) (b) (c)

Figure 3: Multi-temporal analysis of the same slope (camera, year): (a) MOC, 2000; (b) MOC, 2006; (c) HiRISE, 2008.

Since images are acquired by diverse sensors in rather different time periods (of the day and of the season of the year), it is less prone to errors to evaluate the relative difference between the albedo of each streak and its neighbourhood [12-14] rather than computing an absolute albedo value of the streaks in each date. This approach requires that the same and exact streak is correctly identified in each multi-temporal images. The establishment of a correspondence between the same segmented object (with the previous algorithm) in the images along different times is under fine tuning and will be presented elsewhere in a near future.

4 Results

We are currently in a process of building an extensive dataset of annotated images from different locations of Mars captured along different daily periods and solar longitudes so as the diversity of dark slope streaks and different illumination conditions are evaluated. The images consist of the best resolutions available for Mars, that is, those captured by HiRISE (resolutions in the range 0.25-0.50 cm/pixel), by MOC narrow-angle (1.4-6.5 m/pixel) and CTX cameras (6.0 m/pixel). Currently, the dataset is constituted by 210 Martian images of those 3 cameras (Table 1), being the contours of each slope streak delineated manually by an expert (ground-truth).

The performance of the identification algorithm is evaluated through the comparison of each output image with the corresponding ground-truth image by computing the following quantities (pixel-based): precision P(%) = TP/(TP+FP)*100, and an overall quality measure Q(%) = TP/(TP+FP+FN)*100 (TP denotes true positive detections, FP false positive detections and FN false negative detections).

The average results obtained for each sensor and globally are presented in Table 1. They can be considered good enough since the discrepancies between the automated detections and the references are normally on the precise definition of the borders of the streaks and much less on missing/false detections of complete objects. The average quality is a bit higher for HiRISE images than for the other two sensors, we think that is due to the fact that HiRISE images only survey smaller regions on the Martian surface (a consequence of its very high resolution) and normally the scenes under analysis contain mainly slope streaks and very few of other surface features.

Table 1. Average performances of dark slope streaks identification.

Camera	*Images*	P %	Q %
CTX	100	92.7	80.4
HiRISE	20	90.5	83.6
MOC	90	87.8	78.1
Total	**210**	**89.8**	**81.2**

5 Conclusions and on-going developments

These are preliminary but good results based on a single method that has shown a high robustness. The work is still in progress but the detection performances achieved so far can be considered good: overall average quality of 81% in a dataset of 210 images of different sensors and resolutions (from 0.25 to 6.10 m/pixel) from rather different geographic locations of the planet.

Alternative approaches should be tested (*i.e.*, watershed analysing the dynamics of the contours to solve the over-segmentation), also for obtaining a finer segmentation, that is, to separate and characterize adjacent or 'overlapping' streaks with relatively distinct albedo values.

Although the image dataset is already representative of the diversity of the Martian dark slope streaks, its enlargement is still underway, especially in the regions that have a good multi-temporal coverage.

References

.[1] R.Sullivan, P. Thomas P., Veverka J., Malin M., Edgett K. S.,Mass movement slope streaks imaged by the Mars Orbiter Camera. *Journal of Geophysical Research* 106(E10): 23607-23633, 2001.

[2] Schorghofer N., King M.,Sporadic formation of slope streaks on Mars. *Icarus* 216: 159-168, 2011.

[3] Aharonson O., Schorghofer N., Gerstell M.F., Slope streak formation and dust deposition rates on Mars. *Journal of Geophysical Research* 108(E12): 5138, 2003.

[4] Schorghofer N., Aharonson O., Khatiwala S., Slope streaks on Mars: Correlations with surface properties and the potential role of water. *Geophysical Research Letters* 29, 2126: 41-1-41-4, 2002.

[5] Head J.W., Marchant D.R., Dickson J.L., Levy J.S., Morgan G.A., 2007, Slope streaks in the Antarctic Dry Valleys: Characteristics, candidate formation mechanisms, and implications for slope streak formation in the Martian environment. In *LPSC2007 - Lunar and Planetary Science Conference XXXVIII*, Abs. #1935, Houston TX, 2007.

[6] F. Puga., P. Pina, E. A. Silva. Automatic slope streak detection on Mars. In Proceedings of *ISPA2015-9th International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, 2015.

[7] F. Puga., P. Pina, E. A. Silva. Automatic slope streak detection on MOC, CTX and HiRISE images. In *EPSC2015-10th European Planetary Science Congress*, EPSC Abstracts, Vol. 10, EPSC2015-795, 2015.

[8] T. Statella, P. Pina P., E. A. Silva. A study on automatic methods based on mathematical morphology for Martian dust devils tracks detection. In C. San Martin C. and S.-W Kim (eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, LNCS series, vol. 7042: 533–540, Springer-Verlag, Berlin Heidelberg, 2011.

[9] T. Statella, P. Pina P., E. A. Silva. Image processing algorithm for the identification of Martian dust devil tracks in MOC and HiRISE Images. *Planetary and Space Science*, 70: 46–58, 2012.

[10] P. Soille. Morphological Image Analysis. Principles and Applications. 2nd edition. Springer-Verlag, Berlin, 2004.

[11] N. Otsu, A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9: 62–66, 1979.

[12] T. Statella, P. Pina P., E. A. Silva. Automated determination of the orientation of dust devil tracks in Mars orbiter images. *Advances in Space Research*, 53: 1822–1833, 2014.

[13] T. Statella, P. Pina, E. A. Silva. Martian dust devil tracks width measurements. In *EPSC2014-The 9th European Planetary Science Congress*, EPSC Abstracts, Vol. 9, EPSC2014-4, Estoril, Portugal, 2014.

[14] T. Statella, P. Pina P., E. A. Silva. Extensive computation of albedo contrast between Martian dust devil tracks and their neighboring regions. *Icarus* 250: 43–52, 2015.

.

Index of Authors