universidade de aveiro deti departamento de electrónica, telecomunicações e informática ieeta

Proceedings of

RECPAD 2016

22nd CONFERENCE VEIRO

University of Aveiro, October 28th, 2016

APRP

RecPad 2016

22nd Portuguese Conference on
Pattern Recognition

Aveiro, Portugal
October 28th, 2016

Program Committee Chair

Armando J. Pinho

Organizing Committee

Armando J. Pinho
Diogo Pratas
Raquel Sebastião
Samuel Silva
Sónia Gouveia
Susana Brás

Secretariat

Anabela Viegas

Organized by

IEETA — Institute of Electronics and Informatics Engineering of Aveiro
DETI — Department of Electronics, Telecommunications and Informatics
University of Aveiro

In cooperation with

APRP — Associação Portuguesa de Reconhecimento de Padrões

Contents

Preface

After travelling across Portugal, visiting Vila Real in 2010, Porto in 2011, Coimbra in 2012, Lisboa in 2013, Covilhã in 2014 and Faro in 2015, the Portuguese Pattern Recognition community meets once again in Aveiro. Seven years have passed since RecPad 2009, held at Aveiro, and the motivation fostering the organization of the very first event still stands — there is an important research community in Portugal actively working on topics that directly involve or are related to Pattern Recognition. Moreover, this community values the opportunity provided by RecPad to, once a year, gather young and experienced researchers to discuss their work, exchange ideas and learn among peers. The number of attendees (around 60) at this year's event is an evidence of such involvement. Hence, it is our reinforced belief that this one-day, posters-only model of conference, promoted by the APRP — the Portuguese Association for Pattern Recognition —, is clearly the right approach.

In this 22^{nd} edition of the Portuguese Conference on Pattern Recognition, held at the University of Aveiro, on the 28th of October 2016, 42 works were included in the program, from a total of 46 submissions. All submissions were reviewed by at least two members of the Technical Committee and span a rich set of research institutions, technically sound methods and application areas. To highlight the quality of the work carried out by the Pattern Recognition community present at RecPad, the closing session includes the announcement of a Best Paper Award, as voted by the conference attendees.

As in the past editions of RecPad, we continue to include a keynote lecture on a relevant topic to the community, given by an invited speaker. This year, we invited Prof. Joan Serra-Sagristà, from the Universitat Autònoma de Barcelona (UAB), who very kindly accepted to present his work on Remote Sensing Data Compression.

On behalf of the organizing committee, we would like to thank all that were involved in this event, namely, the members of the Technical Committee, for their work in providing constructive feedback to the authors, the Portuguese Association for Pattern Recognition and its president, Prof. Luís Alexandre, for supporting the presence of the invited speaker, and to the University of Aveiro, for giving logistic support to the organization of this conference. A special thanks to Anabela Viegas, for her tireless help, shared experience and invaluable advice.

We sincerely hope that you enjoy this edition of RecPad!

The Organizing Committee

Keynote Lecture

Remote sensing data compression

Joan Serra Sagristá

Summary

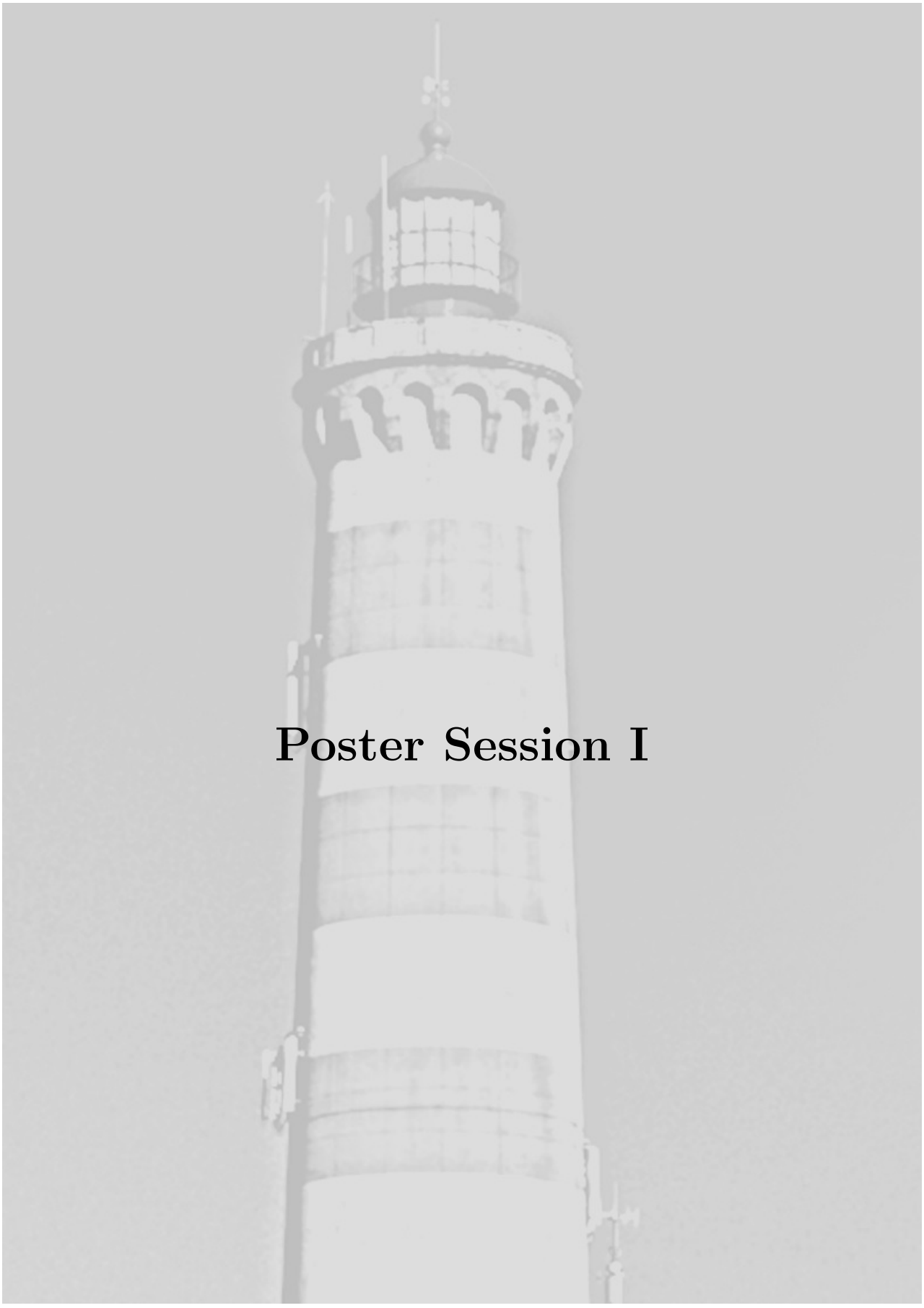
This talk describes recent developments in several areas of remote sensing data compression. The first part of the talk will introduce the current status of Earth Observation missions and the need for efficient data transmission, where data compression plays a significant role. The second part of the talk will be dedicated to onboard compression of remote sensing data, in particular to recent and ongoing work developed by the main space agencies. The third part of the talk will introduce some of our own recent developments in this field.

Short biography

Joan Serra-Sagristà (IEEE Senior Member 2011) received his Ph.D. degree in computer science from Universitat Autònoma de Barcelona (UAB), Spain, in 1999. He is currently an Associate Professor at Department of Information and Communications Engineering, UAB. He holds the Accreditation as Full Professor from both Spanish ANECA and Catalan AQU Catalunya. From September 1997 to December 1998, he was at University of Bonn, Germany, funded by DAAD. His current research interests focus on data compression, with special attention to image coding for remote sensing and telemedicine applications. He serves as Associate Editor of IEEE Trans. on Image Processing and as Program Committee co-chair for IEEE Data Compression Conference. He has co-authored over one hundred publications. He was the recipient of the Spanish Intensification Young Investigator Award in 2006.

Program Committee

Alexandre Bernardino (IST)
Ana Maria Mendonça (FEUP)
Ana Maria Tomé (UA)
António J. R. Neves (UA)
António Pinheiro (UBI)
Armando Pinho (UA)
Aurélio Campilho (FEUP)
Beatriz Sousa Santos (UA)
Bernardete Ribeiro (UC)
Catarina Silva (IPL)
Diogo Pratas (UA)
Fernando Monteiro (IPB)
Hans du Buf (UAlg)
Helder Araújo (UC)
Hugo Proença (UBI)
Jaime Cardoso (FEUP)
João Barroso (UTAD)
João Rodrigues (UAlg)
João Sanches (IST)
João Tavares (FEUP)
Joaquim Pinto da Costa (FCUP)
Jorge Barbosa (FEUP)
Jorge S. Marques (IST)
Jorge Santos (ISEP)
José Silva (Academia Militar)
Luís A. Alexandre (UBI)
Luís F. Teixeira (FEUP)
Mário Figueiredo (IST)
Noel Lopes (IPG)
Paulo Oliveira (IST)
Paulo Salgado (UTAD)
Pedro Pina (IST)
Raquel Sebastião (UA)
Samuel Silva (UA)
Sónia Gouveia (UA)
Susana Brás (UA)
Susana Vinga (IST)
Verónica Vasconcelos (ISEC)



Poster Session I

Why should you model time when you use Markov Models for heart sound analysis

Jorge Oliveira[1]
Theofrastos Mantadelis[2]
Miguel Coimbra[1]

[1]Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto.

[2]CRACS & INESC TEC, Faculdade de Ciências da Universidade do Porto

Abstract

In this paper, we propose a model to segment heart sounds using a semi-hidden Markov model instead of a hidden Markov model. Our model in difference from the state-of-the-art hidden Markov models takes in account the temporal constraints that exist in heart cycles. We experimentally confirm that semi-hidden Markov models are able to recreate the "true" continuous state sequence more accurately than hidden Markov models. We achieved a mean error rate per sample of 0.23.

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 [1]. The signal processing of a PCG has two main goals: the first one, is to split the PCG into heart cycles. Each heart cycle is composed by the first heart sound (S1), the systolic period (siSys), the second heart sound (S2), and the diastolic period (siDia). The second goal is the detection of other sounds such as the third and fourth heart sounds (S3 and S4 respectively) as well as heart murmurs that may be associated to cardiac pathologies.Recently, HMMs has shown to be very effective in modelling the heart sound signals (see Figure 1): Schmidt [2], implemented a duration-dependent HMM using the homomorphic filtering envelogram as observation to the system. This has the advantage (compared to the traditional HMM) that every state duration is explicitly modeled in the state transition matrix. The state duration distribution function is modeled by a Gaussian distribution, where the systolic (siSys) and diastolic (siDia) duration parameters are estimated through autocorrelation analysis of the homomorphic filtering envelogram. Our contributions are: (1) we present an alternative approach for modeling the sojourn time (waiting time) by a semi-hidden Markov Model (HSMM); (2) we approximate the sojourn time distribution by using a Poisson distribution; (3) we conduct experiments over 32 different models (4 suitable observation features ×4 dataset fractions ×2 model types) over a real life dataset of 13 individuals.

2 Hidden-Semi Markov Models

In standard HMMs, the sojourn time (waiting time) is geometrically distributed over all states. This is an unrealistic assumption in heart sound signals, since the state transition probabilities are constantly changing over time. The solution we propose is to model explicitly the sojourn time by using a HSMM.

Let π be the initial state distribution, P a continuous Gaussian state depended distribution matrix, Γ the state transition rules and D the Poisson sojourn time distribution. A HSMM is specified by the quadruple $\Theta = \{\pi, \mathrm{P}, \Gamma, D\}$ [3].

Our goal is to estimate the state-sequence $C_1^T = \{c_1 = S1, \ldots., c_l = S2, \ldots, c_T = Diastoly\}$, which is most likely given set of observations $X_1^T = \{x_1, \ldots. , x_T\}$.

$$S^* = \operatorname{argmax} \Pr(C_1^T | X_1^T, \theta) \quad (1)$$

The parameters Θ are estimated using the expectation maximization (EM) algorithm, which assigns posterior probabilities to each component with respect to each observation. The method uses an iterative algorithm that converges to an optimal solution [3].

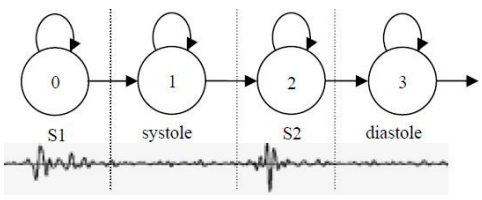


Figure 1: Four state HMM for a cycle of a normal heart sound signal, adapted from [4].

3 Using the Poisson distribution

For our application, the duration probability distribution D is approximated by a Poisson distribution. We chose the Poisson distribution, as the more suitable to model the behavior of PCG signals because:

1) We need to count the number of state transitions in a large number of trials (n-sampling size).

2) The state transitions in PCG signals are rare events because we sample at a high frequency (our signals have a sampling rate f = 4 kHz).

3) The successful events are also very unlikely in heart sound signals, because of some physiological time constraints that exist in the cardiac cycle, for example: the cardiac muscle (like any excitable tissue) exhibits a refractory period to re-stimulation. During this time interval normal cardiac impulse cannot re-excite an already excited area of cardiac muscle. The normal refractory period of the ventricle is 0.25 to 0.30 second [1].

To use a Poisson distribution, we have made an assumption that the outcome trials in different time instances are "weakly" dependent but not necessarily independent.

4 Materials

The dataset we use is composed by samples from 13 healthy individuals from six months to 17 years old (one for each participant). The heart sounds have been collected in Real Hospital Português (Recife, Brasil) using a Littmann 3200 stethoscope embedded with the DigiScope technology. The DigiScope technology was developed within the homonymous project to collect, transmit and record heart sounds without interfering with clinical routine [5]. The heart sounds are recorded in the mitral spot, for about 15 seconds at 4000Hz sampling frequency. Two cardiacpulmonologists manually annotated the S1 and S2 states beginning and ending using the Audacity software.

5 Experimental Setup

In order to optimize the HMM and HSMM parameters we used the EM algorithm [3] also known as the Baum-Welch algorithm. In our experimental setup all states have equal starting probabilities (π). To compute the initial parameters P we use an envelogram segment around the corresponded annotated states. To compute the initial parameters D. we use the annotated time lapse between the beginning and the end of the corresponding states. We use from $\frac{1}{3}^{rd}$ to $\frac{3}{4}^{th}$ of the first part of each subject signal to initialize P, D and use the rest as a test dataset. Since insufficient annotated heart beats are used for computing the initial parameters, the initial standard deviations σ of the Gaussian distributions are biased. In order to reduce the bias, we widen the Gaussian distribution by multiplying σ by a factor of 10^2. We use this methodology because the true parameters diverge from subject to subject. Attempting to generalize over all subjects, leads to inaccurate parameter initialization and the algorithm does not converge to an optimum solution. The signal features are extracted in Matlab and the experiments are conducted using mhsmm package for R [3]. For more implementation details on the Viterbi or the Baum- Welch algorithm see [3].

6 Feature Extraction

The system first filters the original signal using a Butterworth bandpass filter of order 10. We use a lower cutoff frequency of 30Hz and a higher cutoff frequency of 460Hz. From the filtered signal, different envelograms are extracted: Shannon energy in the frequency domain [6]; homomorphic filtering [7]; Shannon energy in the time domain [8]; and the entropy gradient [9]. Shannon energy in the frequency domain which

we compute as in [6] accentuates the pressure differences found across heart valves, which leads to distinct frequency signatures of the valve closing sounds. Homomorphic filtering, the signal is viewed as a product of slowly varying components (heart sounds) with fast oscillatory components (noise). These fast components are rejected by applying a non-linear transformation and are compute as in [7]. Shannon energy in the time domain which is computed as in [8], is used to emphasize the medium intensity of the signals and attenuate high intensity. This tends to make medium and high intensity signals similar in amplitude. Finally, entropy gradient envelogram measures the predictability of the heart sound components in a signal by looking to the total entropy fluctuation in the "expanded region" as the original time series is shifted in a circular motion and is compute as in [9].

7 Results

The performance of the HSMM and HMM was measured as the model's capacity to recreate the continuous state sequence annotated by the cardiacpulmonologists. The mean error rate per sample is calculated by:

$$\bar{\mu}_E = \frac{\#False\ labeled\ Samples}{\#Samples} \quad (2)$$

A sample in the instance t is labeled falsely, when the predicted state of the sample and the annotated state of the sample are not equal. As can be seen from the Figure 2, the best results with the standard HMM has a $\bar{\mu}_E = 0.57$, using the Homomorphic filtering as observational input and a training dataset fraction size of $\frac{3}{4}^{th}$. The best HSMM results were achieved using Shannon energy in the frequency domain with a $\bar{\mu}_E = 0.23$, for a training dataset fraction size of $\frac{1}{3}^{rd}$. For all types of observational input, the HSMM performed better than the equivalent HMM.

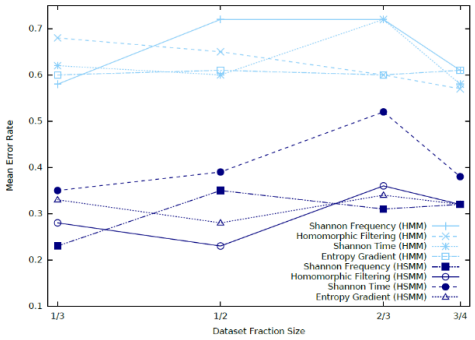
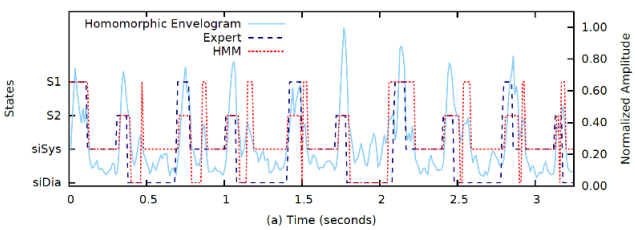


Figure 2: Mean error rate per sample for the hidden Markov model and hidden semi-Markov model.

The HMM is not capable to detect the right sequence of events and not even the state duration in each state as it can be seen in Figure 3(a), this might be a consequence of using static state transition matrix. In the HSMM, the assumption that the Markov chain is homogeneous is dropped, instead it is assumed that the state transition matrix is dependent on time (following a Poisson distribution in the present case), and this ultimately leads to a model capable of describing the non-stationary events in the heart sound signal as depicted in Figure 3(b) with more accuracy than the standard HMM. Finally, we noticed in some case that the starting state was misclassified and as a result the signal has classified reversely by the HSMMs, These signals are properly classified if we set an observed starting state.



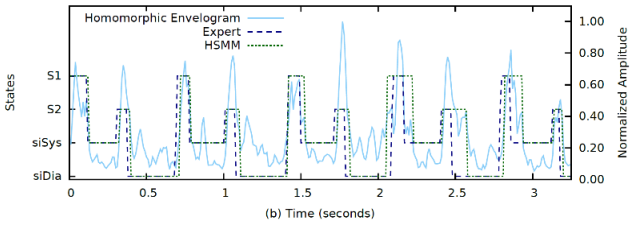


Figure 3: Classification results of heart sound recordings from a normal subject. The dashed lines present the states classified by an expert, a) **(HMM)**, b) **(HSMM)**; and the solid lines present the observation input to the model.

8 Conclusions

In this paper, a heart sound classification algorithm is proposed using HMMs or HSMMs, furthermore we used four different type of features as input to the system. These features are very descriptive and sensitive to S1 and S2 events. Our experiments shows that HSMM outperformed HMM regardless the observational features tested, this suggests that using information concerning the duration probability distribution in each state is a requirement step in modelling heart sound signals. We approximated the duration probability distribution by the Poisson distribution. For future work, we intend to conduct extensive experiments with different distributions in order to approximate the duration probability distribution of the HSMM. Furthermore, we want to conduct experiments where we test the influence of heart rate variability (such as the heart rate of infants and subjects with arrhythmia) in HMMs or HSMMs.

Acknowledgement

This article is a result of the project NanoSTIMA, NORTE-01-0145-FEDER-000016, supported by Norte Portugal Regional Operational Programme (NORTE 2020), through Portugal 2020 and the European Regional Development Fund.

References

[1] J. E. Hall and A. C. Guyton, Textbook of medical physiology. Philadelphia, Pa.: Saunders/Elsevier, 12th ed., 2011.

[2] S. Schmidt, E. Toft, C. Holst-Hansen, C. Graff, and J. Struijk, "Segmentation of heart sound recordings from an electronic stethoscope by a duration dependent hidden-markov model," in Computers in Cardiology, pp. 345–348, 2008.

[3] J. O Connell and S. Højsgaard, "Hidden semi markov models for multiple observation sequences: The mhsmm package for r," Journal of Statistical Software, vol. 39, no. 1, pp. 1–22, 2011. https://cran.rproject.org/web/packages/mhsmm/mhsmm.pdf.

[4] Y.-J. Chung, Pattern Recognition and Image Analysis, Iberian Conference, Classification of Continuous Heart Sound Signals Using the Ergodic Hidden Markov Model, pp. 563–570. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[5] D. Pereira, F. Hedayioglu, R. Correia, T. Silva, I. Dutra, F. Almeida, S. Mattos, and M. Coimbra, "DigiScope - Unobtrusive Collection and annotating of auscultations in real hospital environments," in Engineering in Medicine and Biology Society, IEEE Conference, pp. 1193–1196, 2011.

[6] D. Kumar, P. Carvalho, M. Antunes, R. P. Paiva, and J. Henriques, "Noise detection during heart sound recording using periodicity signatures," Physiological Measurement, vol. 32, no. 5, p. 599, 2011.

[7] C. N. Gupta, R. Palaniappan, S. Swaminathan, and S. M. Krishnan, "Neural network classification of homomorphic segmented heart sounds," Appl. Soft Comput., vol. 7, no. 1, pp. 286–297, 2007.

[8] H. Liang, S. Lukkarinen, and I. Hartimo, "Heart sound segmentation algorithm based on heart sound envelogram," in Computers in Cardiology, pp. 105–108, 1997.

[9] J. Oliveira, A. C. Castro, and M. Coimbra, "Exploring embedding matrices and the entropy gradient for the segmentation of heart sounds in real noisy environments," in Engineering in Medicine and Biology Society, IEEE Conference, vol. 1, pp. 3244–3247, 2014.

Tracking of the Anterior Mitral Leaflet in Echocardiographic Sequences using Active Contours

Malik Saad Sultan[1,2]
engr.saadsultan@gmail.com
Nelson Martins[1,3]
nelsonmartins89@gmail.com
Miguel Tavares Coimbra[1,2]
mtcoimbra@gmail.com

[1] Faculdade de Ciências, Universidade do Porto, Portugal

[2] Instituto de Telecomunicações, Porto, Portugal

[3] Enermeter, Sistemas de Medição, Lda, Braga, Portugal

Abstract

Echocardiography assessment of cardiac valves plays a vital role in the diagnosis of rheumatic heart disease. In the vast majority of cases, the mitral valve gets affected, leading to the thickening of its leaflets that may result in the fusion of their tips. This changes the appearance and reduces the mobility of the leaflets, which also reduce the heart efficiency. Quantifying such parameters provides diagnostic insight. To achieve that, the first step is to identify and then track fast moving leaflets. This work is focused on Anterior Mitral Leaflet (AML) tracking. Open ended active contours are employed in this work by removing its boundary conditions. The external and internal energy of the contour is modified that extend the capture range, improve snake energy and encourages the leftmost end point of the contour to converge on the moving tip of the AML. Results show that contour points are tracked accurately with an average error of 4.9 pixels and a standard deviation of 2.1 pixels in 9 fully annotated normal sequences of real children clinical assessments.

1 Introduction

Rheumatic Heart Disease is one of the serious consequences of Acute Rheumatic Fever. Acute Rheumatic Fever is the inflammatory disease that usually begins in childhood, and whose repeated episodes slowly damage the valves of the heart. Since Rheumatic Heart Disease doesn't occur after the very first attack, the early detection is considered vital to define the disease burden and to control disease progression [1, 2]. Echocardiography can play a key role by providing early evidence, since one can confirm the suspected cases (valve involvement), which can be treated accordingly [3, 4]. The Paresternal Long Axis (PLA) view is the most suitable view that allows us to access the mitral valve [5] (Fig. 1).

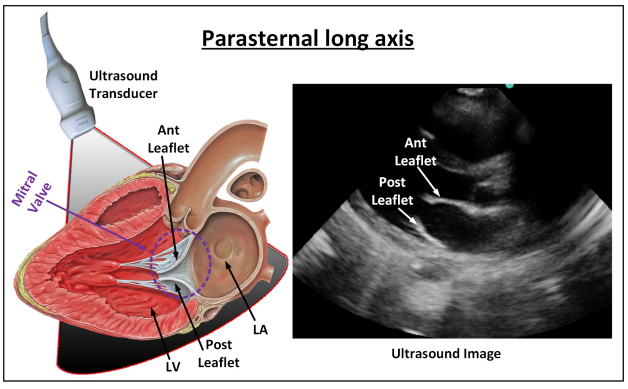


Figure 1: Parasternal long axis view of a normal mitral valve in diastole.

Active contours [6] were widely used in segmentation applications, however several shortcomings of the algorithm were found such as, initialization, convergence, etc. In this work, the classical snake model is modified to track thin and elongated anterior mitral leaflets in the parasternal long axis view of ultrasound images. Open ended contours were used and external energy of the model is adapted in a way that increases the range of snake and encourages the end point of the snake to stay on the tip of the anterior mitral leaflet.

2 Classical Snake

The classical snake model considers a parameterized curve that evolves in the spatial domain. Since the proposed work deal with images, the parameterized curve has been discretized as N sample/control points. The total energy of the classical snake consists of external and internal energy (1).

$$\int_0^1 \left(\overbrace{\frac{1}{2}\left[\alpha \underbrace{\left\| \frac{\partial V}{\partial S} \right\|^2}_{Elasticity} + \beta \underbrace{\left\| \frac{\partial^2 V}{\partial S^2} \right\|^2}_{Stiffness} \right]}^{Internal} + \overbrace{E_{ext}(V)}^{External} \right) ds \tag{1}$$

3 Improved Snake Model

The proposed method will focus on tracking alone and so it assumes perfect segmentation of the AML in the first frame.

3.1 Internal Energy

The internal energy is responsible to regulate the contour by imposing the elasticity and stiffness that are the first and second derivative, respectively. The weights, α and β controls the relative influence of each term (1). The physical length of the AML does not change in the PLA view, this measure of internal energy is not adequate for our problem. Only stiffness is used that controls the bending energy of the snake. The weight β remains constant for all the contour points. Free boundary condition is used that allows the end points of the contour to move freely on the image plane to get minimum energy, promoting line contours instead of circles. However, it is bounded to remain closer to its neighbour contour point (2).

$$x^{t+1} = (M+\gamma I)^{-1} x^t, y^{t+1} = (M+\gamma I)^{-1} y^t$$

$$\underbrace{\begin{bmatrix} \hat{r} & q & p & & & & \\ \hat{q} & r & q & p & & & \\ p & q & r & q & p & & \\ & \ddots & \ddots & \ddots & & & \\ & & p & q & r & q & p \\ & & & p & q & r & \hat{q} \\ & & & & p & q & \hat{r} \end{bmatrix}}_{M} \begin{cases} p = \beta \\ q = -4\beta \\ r = 6\beta \\ \hat{r} = 3\beta \\ \hat{q} = -3\beta \end{cases} \tag{2}$$

Whereas, γ control the step size and is always positive and I is an identity matrix.

3.2 External energy

In this work external image energy of the snake is modified so that it best fits with our particular application. Instead of using a simple intensity image and the edge map, we use Difference of Gaussian (DoG), which is a simplified way to approximate the Laplacian of Gaussian. DoG was found very effective in discarding high frequency details from ultrasound images [7]. DoG is divided into two main parts: pixels with positive values were characterized as high-intensity region and negative valued pixels characterized as edges (3) (Fig. 2).

$$\begin{aligned} E_{DoG\_line} &= I \times (G_{\sigma_1} - G_{\sigma_2}) \quad DoG \geq 0 \\ E_{DoG\_edge} &= I \times (G_{\sigma_1} - G_{\sigma_2}) \quad DoG < 0 \end{aligned} \tag{3}$$

In the above equation, the image is convolved with the difference of Gaussian kernels of width σ_1 and σ_2 whereas, $\sigma_2 > \sigma_1$. In this work, we used the Gaussian filter of size 20 with sigma 8 and 10, obtained empirically in pilot tests.

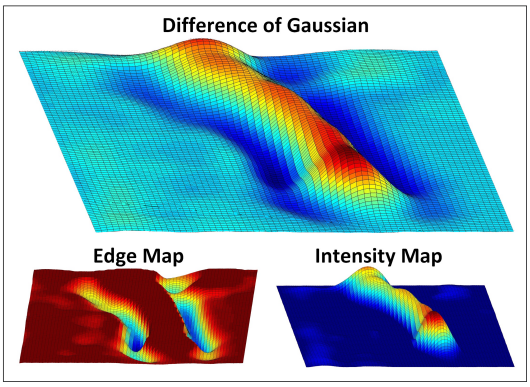


Figure 2: Surf plots: DoG of AML region of grayscale image. Positive valued pixels (intensity map), negative valued pixels (edge map).

To obtain better tracking performance, end points of the contour are encouraged to continuously follow the tip of the AML. To achieve this, cornerness energy is used, instead of line or edge energy, that encourages end points to stay on the AML tip. The methods based on auto-correlation were used in this work. Moravec [8] had explored that the maximum intensity variation in various directions within the local window represents a corner. Later on, Harris et al. [9] has improved his approach by proposing analytic expansion of the shift (4) (Fig. 3)

$$S = \sum_x \sum_y \underbrace{w(x,y)}_{window} \left(\underbrace{I(x+u, y+v)}_{shifted\_intensity} - \underbrace{I(x,y)}_{intensity} \right)^2 \tag{4}$$

Window function can be a Gaussian or rectangle at position x, y. (u, v) shows a small displacement in all directions that somehow estimate the intensity difference. The modified edge and intensity snake energy provide an attraction force that attracts all the points except the ones close to the AML tip.

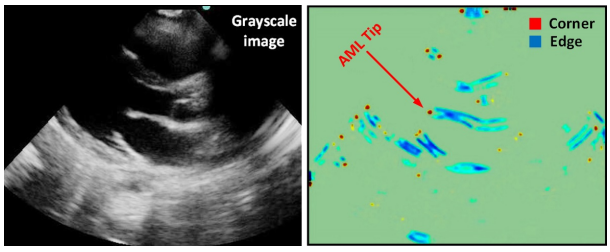


Figure 3: Harris Cornerness measure (E_{Harris}), shows corners in Red and edges in Blue

4 Results

4.1 Materials

In one of the activities of Real Hospital Português, in Recife, Brazil, a dataset of ultrasound mitral valve videos has been collected for the purposes of screening acute rheumatic fever in children. The data was collected using a M-Turbo model by SonoSite ultrasound system, with C11x transducer (5-8 MHz). Nine of these exams were fully annotated (manual segmentation of all frames) using support software and were used to test the novel algorithm proposed in this work. These nine videos include a total of 1137 frames with dimensions of $[351 \times 441]$.

4.2 Tracking results

The proposed algorithm used the manual annotation as its contour initialization for the first frame of every video sequence. The algorithm manages to handle small to medium frame to frame displacement. However, the algorithm fails to track if the displacement is very high (about 33 pixel). The Hausdorff distance error between the set of manually annotated points (GS) and snake control points (SEG) is used.

The algorithm is also compared with the classical snake approach exhibiting superior performance (table. 1).

Table 1: Tracking error (Pixel)

Video No.	***No. Of Frames***	Our Approach	Kass Approach
		Avg / STD *Error (PX)*	***Avg / STD*** *Error (PX)*
1	131	5.32 / 1.89	6.08 / 3.26
2	360	4.6 / 1.8	5.7 / 1.85
3	66	5.22 / 2.4	7.84 / 2.47
4	131	4.33 / 1.81	6.7 / 1.67
5	66	5.6 / 4.06	7.66 / 3.95
6	66	5.74 / 1.56	5.15 / 2.19
7	120	4.95 / 1.96	5.61 / 1.66
8	66	4.97 / 2.27	6.1 / 1.94
9	131	3.67 / 1.72	5.56 / 2.45
Total	**1137**	**4.93 / 2.16**	**6.26 / 2.38**

Acknowledgement

This article is a result of the project (NORTE-01-0247-FEDER-003507-RHDecho), co-funded by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This work also had the collaboration of the Fundação para a Ciência a e Tecnologia (FCT) grant no: PD/BD/105761/2014 and has contributions from the project NanoSTIMA , NORTE-01-0145-FEDER-000016, supported by Norte Portugal Regional Operational Programme (NORTE 2020), through Portugal 2020 and the European Regional Development Fund (ERDF).

References

[1] A. Bisno, E. G. Butchart and et al. "Rheumatic fever and rheumatic heart disease," Who Tech. Rep. Ser., pp. 923-1122, Nov. 2001.

[2] J. R. Carapetis "The stark reality of rheumatic heart disease," European Heart Journal, vol. 36, no. 18, pp. 1070-1073, May 2015

[3] E. Marijon, D.S. Celermajer and et al. "Rheumatic heart disease screening by echocardiography: the inadequacy of World Health Organization criteria for optimizing the diagnosis of subclinical disease," Circulation, vol. 120, no. 8, pp. 663-668, Aug 2009

[4] B. Remenyi, N. Wilson and et al. "World heart federation criteria for echocardiographic diagnosis of rheumatic heart disease an evidence-based guideline," Nat. Rev. Cardiol., vol. 9, no. 5, pp. 297-309, Feb. 2012

[5] A.S. Omran, A.A. Arifi, A.A. Mohamed, "Echocardiography of the mitral valve," Journal of the Saudi Heart Association, vol. 22, no. 3, pp. 165-170, Feb. 2010

[6] M. Kass, A. Witkin, D. Terzopoulouse, "Snakes : active contour model," Int'l journal of Computer Vision, vol. 1, no. 4, pp. 321-331, Jan. 1988

[7] M. W. Davidson, M. Abramowitz "Molecular Expressions Microscopy Primer: Digital Image Processing Difference of Gaussians Edge Enhancement Algorithm," Olympus America Inc., and Florida State University

[8] H.P. Moravec, "Towards automatic visual obstacle avoidance," 5th Int'l Joint Conf. On Artifcial Intelligent, pp. 584, 1977

[9] C. Harris, M. Stephens, "A combined corner and edge detector," Alvey vision Conf., 1988

Pedestrian Detection Using Multi-Stage Features in Fast R-CNN

Miguel Farrajota
mafarrajota@ualg.pt
J.M.F. Rodrigues
jrodrig@ualg.pt
J.M.H. du Buf
dubuf@ualg.pt

Vision Laboratory, ISR (Lisbon), LARSyS,
University of the Algarve
Campus de Gambelas, 8005-139 Faro, Portugal

Abstract

Pedestrian detection in general is a difficult task due to a large variability of features due to different scales, views and occlusion. Typically, smaller and occluded pedestrians are hard to detect due to fewer discriminative features if compared to large-size, visible pedestrians. In order to overcome this we use convolutional features from different stages in a deep Convolutional Neural Network (CNN), with the idea of combining more global features with finer details. In this paper we present an object detection framework based on multi-stage convolutional features for pedestrian detection. This framework extends the Fast R-CNN framework for the combination of several convolutional features from different stages of the used CNN to improve the network's detection accuracy. The Caltech Pedestrian dataset was used to train and evaluate our method.

1 Introduction

Detecting pedestrians by identifying visible persons is difficult because of variations in the target appearance, pose, size, lighting and occlusion. Moreover, each independent variation affects detection differently, but the two main effects that hamper detection most are scale and occlusion [7]. Existing work has tackled the scale variation problem in several ways. Data augmentation techniques [4] like resizing and multiple scales have been used to increase robustness to scale variations. Other methods used a single model but with several filters tuned to specific scales which are applied to all pedestrians with various sizes. This, however, cannot solve the problems due to the large intra-class variation of small and large persons. Li et al. [7] combined a large-size sub-network with a small-size one for detecting pedestrians of varying sizes. The use of a weighted score of both sub-network responses significantly increased accuracy because each network is tuned to different features.

In this paper we pursue a different strategy in order to cope with feature differences due to person sizes. We present an object detection framework which uses multi-stage features of a deep Convolutional Neural Network (CNN) to improve detection accuracy. By using feature maps from different convolutional layers with different receptive field sizes, we can cope with some ambiguity in discerning pedestrians from background due to the size variability. Since the size of a receptive field depends on the depth of its layer in the network, different fields will code different features of differently sized pedestrians. The proposed method extends the Fast R-CNN [4] framework by using and combining multiple feature maps from different stages of a CNN for classification. The Caltech pedestrian dataset will be used to train and test the proposed method.

The main contribution of this paper is the integration of multiple features from different stages of a deep CNN to improve detection accuracy. While most detection methods do not take advantage of more information available in the CNN pipeline, here we investigate the usefulness of employing more feature maps besides the last convolutional layer, which holds more complex features than the previous layers.

2 Pedestrian detection

The proposed method, Multi-Stage Feature (MSF) Fast R-CNN, is capable of integrating several feature maps from multiple convolutional layers of a CNN and to combine them in a single network of fully-connected layers for classification. It works as follows: the model takes images and a number of Regions-of-Interest (RoI) proposals as input and then outputs detection results. The model is composed of three main components: i) a CNN to extract feature maps from convolutional layers; ii) a RoI pooling layer that extracts sub-sections defined by the input RoI proposal coordinates in the image from two convolutional feature maps in the CNN pipeline; and iii) a final network classifies the extracted sub-sections (pedestrian or background class) and it also outputs refined bounding-box positions.

We used a VGG16 [10] model for feature extraction, which has been trained on Imagenet [9]. We used all layers up to the last max-pooling layer, and during training the first four convolutional layers in the network had their parameters fixed (i.e., they were not optimized). Furthermore, we extracted feature maps from two convolutional layers in the CNN pipeline, from layer 13 which is the last CNN layer and from layer 10. The RoI pooling layers extract feature maps for each RoI proposal with a fixed resolution of 7×7 pixels.

To generate RoI proposals we used the ACF [2] and LDCF [8] detectors. These detectors are publicly available and both use a fast sliding window strategy that performs quite well for rigid object detection. Also, they can be trained to detect specific object categories like pedestrians. This allows us to generate high quality RoI proposals quickly and efficiently. We use the Caltech dataset [1] for training the ACF pedestrian detector, and the generated proposals are then used as input to train our Fast R-CNN network. For evaluation we use a pre-trained LDCF detector to generate proposals for test images because of its smaller miss rate.

The proposed method is trained and evaluated by using the Caltech Pedestrian dataset [1]. This dataset and its benchmark is one of the most popular and challenging publicly available datasets. We used the standard "Reasonable" train/test setting and sampled every other 3rd frame from the set's videos, resulting in about 20k annotated pedestrians extracted from 42782 frames. We followed the proposed evaluation protocol by measuring the log average miss rate over nine points ranging from 10^{-2} to 10^{0} False-Positives-Per-Image (FPPI). We also compare performance with the best-performing methods as suggested by the Caltech benchmark on the "reasonable" subsets, where pedestrians have a height of at least 50 pixels and are occluded at most 65%.

The network was trained using stochastic gradient descent with a momentum of 0.9 and a weight decay of 0.0005. All network weights which were not pre-trained on Imagenet were randomly initialized with a uniform distribution. We used mini-batches of 128 randomly sampled object proposals from two images; 25% of these were positive RoI proposals having an intersect-over-union (IoU) of at least 0.5 with ground truth boxes. The remaining samples were negative object proposals; 25% of these were RoI proposals having an IoU with the ground truth box in the interval [0.1, 0.5), and the remaining 75% of the object proposals had 0% overlap with ground truth boxes. Dropout with 50% chance was applied to all fully-connected layers of the classifier except the first one, and batch normalization [6] was used for faster convergence during training. We updated the network parameters with a learning rate of 0.001 for 4 epochs and then reduced it by 1/10 for an extra 3 epochs, with a total of 7 epochs for training using the same combined loss as in [4]. During training and testing, the scale of the input image was set to 800 pixels on the shortest side. For data augmentation, images were horizontally flipped with a probability of 50%.

3 Discussion and results

We presented a method for pedestrian detection which is based on deep neural networks with multi-stage feature combination. The proposed metho employs multiple convolutional features from different processing layers. This results in an increased detection performance without many extra computations. We demonstrated that combining features from an early stage with those from a later one makes it easier to distinguish pedestrians from background. Also, this combination of global features with finer details performs best when they are fed into a couple of networks that are combined in a final stage of the model.

Results of the method demonstrate the applicability and usefulness of the detector. Figure 1 shows some detection results on the Caltech test dataset [3]. Figure 2 shows benchmark results on the test set with our method and other top-performing methods. Our model shows competitive

Figure 1: Network's architecture illustration (top) and detection results on the Caltech dataset [1] (bottom).

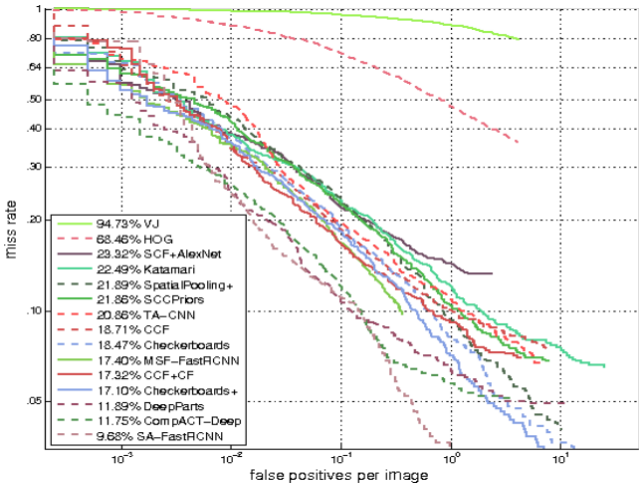


Figure 2: Performance comparison of our method with other state-of-the-art methods on the Caltech dataset [1] (lower is better).

results, placed among the top-10 best performing algorithms, ranking 6^{th} with a 17.40% miss rate (lower is better).

In future work we expect to benchmark the current framework with the most popular CNNs like GoogleNet [11] or ResNet [5], and investigate more advanced region proposal generators besides ACF and LDCF.

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] Piotr Dollár, Christian Wojek, Bernt Schiele, and Pietro Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. PAMI*, 34(4):743–761, 2012.

[2] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. Fast feature pyramids for object detection. *IEEE Trans. PAMI*, 36(8): 1532–1545, 2014.

[3] 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.

[4] Ross Girshick. Fast r-cnn. In *IEEE Proc. ICCV*, pages 1440–1448, 2015.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[6] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[7] Jianan Li, Xiaodan Liang, ShengMei Shen, Tingfa Xu, and Shuicheng Yan. Scale-aware fast r-cnn for pedestrian detection. *arXiv preprint arXiv:1510.08160*, 2015.

[8] Woonhyun Nam, Piotr Dollár, and Joon Hee Han. Local decorrelation for improved pedestrian detection. In *NIPS*, pages 424–432, 2014.

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *IJCV*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[11] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, June 2015.

Benchmark of Swarm-Based Image Segmentation Methods for Medical Imaging

Ana Filipa Rodrigues [1], Carmen Nunes [1]
{a21230206, a21230205} @alunos.isec.pt
Verónica Vasconcelos[2,3]
veronica @isec.pt
Micael Couceiro[4]
micael@ingeniarius.pt

[1] Biomedical Engineering Students, Coimbra Institute of Engineering, Polytechnic Institute of Coimbra
[2] Coimbra Institute of Engineering, Polytechnic Institute of Coimbra, and INESC TEC, OPorto
[3] Coimbra Institute of Engineering, Polytechnic Institute of Coimbra
[4] Ingeniarius, Ltd, and Institute of Systems and Robotics, Coimbra

Abstract

Image segmentation is the process of subdividing the image into its constituent parts. For many applications, segmentation has the purpose of finding an object in an image. Otsu's based image thresholding is a method that returns the optimal threshold of a given image by maximizing the between-class variance function, and is a type of global thresholding in which it depends only in the gray value of the image. Segmentation methods based on swarm optimization techniques have been quite attractive in these tasks and are studied here as a way to check their applicability in medical imaging. This paper consists in the study, implementation and comparison of different segmentation techniques based on the Otsu method, including swarm optimization. To that end, the Otsu's function is adopted to assess the efficiency of the different swarm approaches.

1 Introduction

Image segmentation is the process of partitioning a digital image into multiple regions or objects, *i.e.*, a label is assigned to each pixel in the image, such that pixels with the same label share certain visual characteristics [1]. Multilevel segmentation techniques provide an efficient way to perform image analysis by separating foreground (*i.e.*, objects of interest) from background [2]. Image segmentation is classified into four specific groups, such as histogram thresholding, texture analysis, clustering-based methods, and region-based methods [3].

This section starts by summarizing the Otsu's method, which falls within the category of histogram thresholding. Otsu's based image thresholding was initially proposed back in 1979 [4]. This method returns the optimal threshold of a given image by maximizing the between-class variance function, and is a type of global thresholding that depends only on the gray value of the image. Among all segmentation methods, Otsu's is one of the most successful methods for image thresholding [1]. Nevertheless, solving Otsu's function, which relies on maximizing the between-class variance of a given image, may be a complex optimization problem.

In spite of this, this work presents a benchmark which compares multiple swarm-based optimization algorithms designed to maximize the between-class variance inherent to the Otsu's method. The paper is organized as follows: Section 2 gives an overview of the proposed methodology, describing the swarm optimization methods chosen to optimize Otsu's function. Section 3 describes the dataset used in this paper and the obtained results. The paper is concluded in Section 4.

2 Methodology

In this paper four segmentation methods, based on Otsu's thresholding, are benchmarked: PSO (Particle Swarm Optimization), DPSO (Darwinian Particle Swarm Optimization), FODPSO (Fractional-Order Darwinian Particle Swarm Optimization), and CSO (Chicken Swarm Optimization).

The most basic thresholding method is to choose a fixed threshold value and compare each pixel to that value. However, fixed thresholding often fails if the illumination varies spatially in the image. In order to account for variations in illumination, the common solution to be adopted is an adaptive thresholding. The main difference over the fixed thresholding is that a different threshold value is computed for each pixel in the image [5].

The original Particle Swarm Optimization (PSO) algorithm was developed by Eberhart and Kennedy in 1995 [6]. This method takes advantage of the swarm intelligence concept. The candidate solutions, called particles, travel through the search space to find an optimal solution, by interacting and sharing information with neighbor particles, namely their individual best solution (local best), thus computing the neighborhood best. In each phase of the procedure, the global best solution obtained is always updated in the entire swarm [7].

A general problem with the PSO and similar optimization algorithms is that they may get trapped in local optimum points, the DPSO was created to solve this problem. The Darwinian Particle Swarm Optimization (DPSO) was formulated in search of a better model of natural selection using the PSO algorithm [8], in which many swarms of test solutions may exist at any time. To analyze the general state of each swarm, the fitness of all particles is evaluated and the neighborhood and individual best positions of each of the particles are updated. When a new global solution is found, a new particle is spawned. A particle is deleted if the swarm fails to find a more appropriate state in a defined number of steps [2].

The Fractional-Order Darwinian Particle Swarm Optimization (FODPSO) presented in [7] is an extension of the DPSO, which uses fractional calculus to control the convergence rate of the algorithm [7]. This method is, in simple terms, the same as having multiple PSOs, where particles attempt to find the best solution for their own "survival", with the perk of intrinsically having a memory of past decisions. It benefits from a cooperation paradigm in which particles within each swarm cooperate with one another, while multiple swarms compete to find the most adequate solution, i.e., the optimal solution [9].

At last, the Chicken Swarm Optimization (CSO) can efficiently extract the chickens swarm intelligence to optimize problems. The roosters with better fitness values (*e.g.*, higher between-class variance) have priority for "food access" over the ones with worse fitness values. Moreover, they also search within a wider range of possibilities (*e.g.*, combination of thresholds) than that of the roosters. As for the hens, they can follow their groupmate roosters in the same "search for food" [10]. In this case, this behavior is going to be applied to image segmentation.

2.1 Image Thresholding

Let the term L be intensity levels, for example, one color component for grayscale images and these levels are in the range $\{0,1,2,\ldots, L\text{-}1\}$. Then, the probability distribution p_i can be defined as:

$$p_i = \frac{h_i}{N} \qquad \sum_{i=0}^{L-1} p_i = 1 \tag{1}$$

where i is a specific intensity level in the range $\{0 \leq i \leq L\text{-}1\}$, N is the total number of pixels in the image, and h_i is the number of pixels for the corresponding intensity level I.

The total mean of the image is calculated as:

$$\mu_T = \sum_{i=0}^{L-1} i p_i \tag{2}$$

The m-level thresholding presents m-1 threshold levels t_j, where j=1,2,…., m-1, and the operation is performed as:

$$F(x,y) = \begin{cases} 0, & f(x,y) \leq t_1 \\ \frac{1}{2}(t_1 + t_2), & t_1 < f(x,y) \leq t_2 \\ \vdots & \vdots \\ \frac{1}{2}(t_{m-2} + t_{m-1}), & t_{m-2} < f(x,y) \leq t_{m-1} \\ L-1, & f(x,y) \leq t_{m-1} \end{cases} \tag{3}$$

wherein x and y are the width(W) and height(H), in pixels, of the image of size H × W denoted by $f(x,y)$ with L intensity levels.

In this situation, the pixels of a given image will be divided into *n* classes $D_1,\ldots,D_m$, which may represent multiple objects or even specific features on such objects. The probabilities of occurrence w_j of classes $D_1,\ldots,D_m$ are given by:

$$W_J = \begin{cases} \sum_{i=0}^{t_j} p_i, & j = 1 \\ \sum_{i=t_j-1}^{t_j} + 1 p_i, & 1 < j < m \\ \sum_{i=t_j-1}^{L-1} + 1\ p_i, & j = m \end{cases} \quad (4)$$

The mean of each class μ_j can then be calculated as:

$$\mu_j = \begin{cases} \sum_{i=0}^{t_j} \frac{p_i}{w_j\ q}, & j = 1 \\ \sum_{i=t_j-1}^{t_j} + 1 \frac{p_i}{w_j}, & 1 < j < m \\ \sum_{i=t_j-1}^{L-1} + 1 \frac{p_i}{w_j}, & j = m \end{cases} \quad (5)$$

At last, Otsu's between-class variance can be defined as:

$$\sigma_B = \sum_{j=1}^{m} w_j \left(\mu_j - \mu_T\right)^2 \quad (6)$$

Where w_j is the probability of occurrence. The *m*-level thresholding is reduced to an optimization problem to search for t_j , that maximizes the objective function (J_{max}) of the image being defined as:

$$\varphi = \max_{1<t_i<,\ldots,L-1} \sigma_B\left(t_j\right) \quad (7)$$

The computation of this optimization problem will lead to a much larger computational effort when the number of threshold levels increases [10]. The efficiency of each particle is evaluated with respect to the between-class variance σ_B of the image-intensity distributions estimated by Eq.(6).

3 Results

In this work, we used a set of 11 images of skin lesions, provided by the Hospital Pedro Hispano and the Faculty of Sciences of University of Porto, which have a resolution ranging between 640 × 481 and 768 × 577 pixels. Experiments were conducted for 30 trials for each method, 100 iterations each. For these experiments, the chosen datasets were skin lesions images, benign and malignant (figure 1). Then the image is segmented with different gray levels, in this study were used two, four and six levels. A comparison between methods was made not only visually but also by their fitness value and by the time it took to perform each iteration. Above (figure 2) is an example of a segmentation for one of the skin lesions used in these experiences followed by the graphs (figure 3, figure 4, figure 5) of Fitness VS Time (seconds).

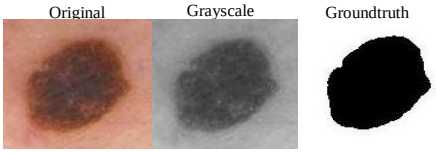


Figure 1- The skin lesion segmented

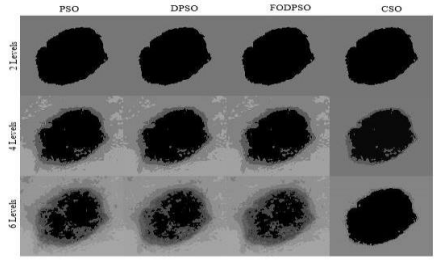


Figure 2- Results for each method and level

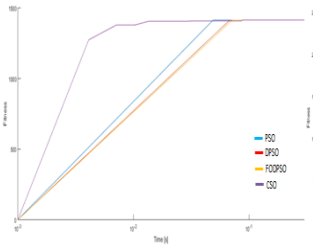


Figure 3- Fitness VS Time graphic, for two levels

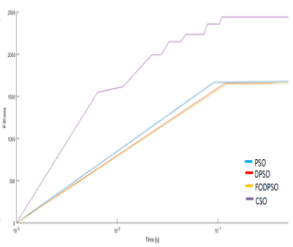


Figure 4- Fitness VS Time graphic, for four levels

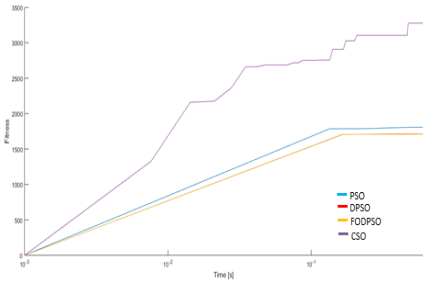


Figure 5- Fitness VS Time graphic, for six levels

4 Conclusion

In this work, four swarm optimization methods were compared with the common goal of maximizing the variance between classes (Otsu function). Once analysed the results, we can conclude that for 2 levels all the methods have very similar results. For 4 and 6 levels, CSO is clearly the best method, although very recent and still little known, it's the one with better results, faster processing time, and better fitness value. For the remaining, PSO, DPSO, and FODPSO present similar results. With this in mind, as future work, a combined architecture will be proposed to merge the outcome from multiple methods, including the ones herein presented, with the intent to provide a higher image segmentation performance for medical imaging applications.

References

[1] H. J. Vala, and A. Baxi. A review on Otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering & Technology*, *2*(2), 387-389, 2013.

[2] P. Ghamisi, M. S. Couceiro, F. M. Martins, and J. Atli Benediktsson. Multilevel image segmentation based on fractional-order Darwinian particle swarm optimization. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(5), 2382-2394, 2014.

[3] V. Rajinikanth, and M. S. Couceiro. Multilevel Segmentation of Color Image using Lévy driven BFO Algorithm. *Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing* (pp. 1-19). Amritapuri, India: ACM, 2014.

[4] C. H. Bindu. An improved medical image segmentation algorithm using Otsu method. *International Journal of Recent Trends in Engineering*, *2*(3), 88-90, 2009.

[5] D. Bradley, and G. Roth. "Adaptive thresholding using the integral image. *Journal Graphics, gpu, and game tools*, *12*(2), 13-21, 2007.

[6] R. C. Eberhart, and J. Kennedy. A new optimizer using particle swarm theory. *Proceedings of the sixth international symposium on micro machine and human science*, *1*, 39-43, 1995.

[7] P. Ghamisi, M. S. Couceiro, J. A. Benediktsson, and N. M. Ferreira. An efficient method for segmentation of images based on fractional calculus and natural selection. *Expert Systems with Applications*, *39*(16), 12407-12417, 2012.

[8] J. Tillett, T. Rao, F. Sahin, and R. Rao. Darwinian particle swarm optimization. *Proceedings of the 2nd Indian International Conference on Artificial Intelligence* , (pp. 1474-1487), 2005.

[9] P. Ghamisi, M. S. Couceiro, and J. A. Benediktsson. A novel feature selection approach based on FODPSO and SVM. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(5), 2935-2947, 2015.

[10] X. Meng, Y. Liu, X. Gao, and H. Zhang. A new bio-inspired algorithm: chicken swarm optimization. *Advances in swarm intelligence*, 86-94, 2014.

Multi-modal Image Registration for Generation of Complete 3D Models of the Breast: A Technical Review

Sílvia Bessa
silvia.n.bessa@inesctec.pt
Jaime S. Cardoso
jaime.cardoso@inesctec.pt
Hélder P. Oliveira
helder.f.oliveira@inesctec.pt

INESC TEC
Porto, Portugal

Abstract

Image registration is an important research topic in medical imaging, with applications such as computer aided diagnosis and surgery planning. However, medical images are usually obtained from different modalities and describe anatomical structures that are deformed during acquisition, which makes the registration task challenging. Many solutions have been proposed, but the maturity of these algorithms still remains an open problem. This is the case of breast imaging registration. Breast image registration is a key task in creating complete 3D models of the breast, which combines multi-modal images, but few works have been published regarding the field of matching surface and interior radiological information of the breast. This is a drawback that limits the development of tools for planning breast cancer surgeries and predict breast deformities caused by cancer treatment. In this paper, some breast image registration techniques and approaches are described.

1 Introduction

Breast cancer is a public health disease affecting over 1.6 millions of women every year. In Portugal, every day 11 new cases are detected and another 4 women die. While a few decades ago the primary goal of breast cancer treatments was to eliminate cancer, with newer techniques, the aesthetic results now play a special role in the treatment decision process, and an increasing number of women have to live with the consequences of treatments for many years [13]. The involvement of women in the treatment decision process has been proven benefit to accept the resulting outcomes, highlighting the necessity of creating tools that predict the outcomes of each possible option, providing patients with visual clues of the expected results for more conscientious decisions. To develop a breast surgery planning tool, it is necessary to create complete 3D models of the patient's breasts. Attempts to model the breast include the use of Magnetic Resonance Image (MRI), Computed Tomography (CT), Ultrasound (US) and 3D models. But in order to obtain a complete 3D model of the breast for planning surgery interventions, the complementary information of different image modalities has to be combined. Breast images are often acquired with different views, modalities or at different times, which makes image registration an important step to convey multiple and complementary information into a single coordinate system to ease the understand and analysis of data [2]. In the next sections, some prominent breast image registration techniques are described, which are used to combine multi-modal radiological exams or reconstruct surface models of the breast. The main goal of this paper is to provide knowledge about the main advances in breast image registration techniques, while highlighting some limitations that have to be overcome to properly create complete 3D models of the breast.

2 Multi-modal Breast Image Registration

Several methods have been proposed to solve the problem of image registration, but this task is particularly hard for breast data, due to the inhomogeneous, anisotropic nature of the soft-tissue within the breast, and its inherent non-rigidity characteristics [2]. The success of registration methodologies depends on the choice of the geometric transformation, which highly depends on the nature of the data to be registered. Registration techniques can use rigid or nonrigid transformations, but most medical image registration approaches are based on the latter, given the deformable nature of most of the anatomical parts of the human body.

Research in the field of breast image registration has been primarily focused on combining mono and multi-modal radiological images, but few attention has been given to the task of matching surface and interior radiological information of the breast. Combining this information would lead to the generation of patient-specific 3D models of the women breast that can be used in visualization and planning of breast cancer surgeries. The registration of multi-modal radiological exams allows the best characterization of the tumour (location, size and volume) and other characteristics as the glandular density, which combined with the 3D external model of the breast, results in a complete model of the breast, useful to predict the risk of deformity and quantify the aesthetic result after surgical removal of the tumour. However, considering that interior and surface data of the breast are acquired in different poses (prone position for interior data, and upright position for surface data acquisition), it is necessary to determine the transformation between the two models. Despite the developments in multi-modal breast image registration algorithms, there is gap for algorithms that match interior and surface data of breast.

3 3D Interior Model of the Breast

Early detection of breast cancer is key to its successful treatment and improvement of survival rates. Consequently, routine mammogram screenings are recommended for a large percentage of female populations, which justifies the large number of breast image registration techniques focused on X-ray mammogram alignment, either for combining bilateral or temporal images. These strategies provide aid to better visualization of breast lesions, as well as improvements in the detection and diagnosis rates of computer assisted diagnosis systems.

However, different breast imaging modalities bring complementary information that can be advantageously used for these tasks. More recently, strategies have been developed which focus on multi-modal registration of breast images. These methodologies usually map the information of MRI, a 3D and nearly undeformable information, to mammograms, and the dual information is subsequently used in the general pipeline of detection and diagnosis of breast lesions. This mapping is accomplished using transformation models based on synthetic deformations [5, 10], or finding corresponding control points in images, such as anatomical regions as breast boundaries, nipple and pectoral muscle. These control points are then aligned either by iterative optimization [7], or by direct computation of the transformation function [4]. Additionally, deformable finite element methods (FEM) have been specially investigated as physical models to register intra and inter-modality images and model the interior of the breast [4, 5, 7, 10], but these methods present high computational cost. Therefore, alternatives based on parametric models have been explored to reduce the complexity of the models to be registered, while preserving the ability to properly align them. Examples of this strategy include the use of Non Uniform Rational Basis Spline (NURBS) [1, 11], or a combination with Free Form Deformation (FFD) to obtain deformable models of the breast [3, 9].

In spite of the encouraging results accomplished by combining multi-modal breast images in the detection and diagnosis of breast lesions, there is no follow up of this task to the planning of breast cancer surgery. Surgery planning still relies on manual drawings, and rude marks drawn in the patient's body, and would benefit from the inclusion of image registration outputs in a model suitable for visualization.

4 Surface Model of the Breast

To have a complete representation of the breast, it is necessary to obtain a surface model of the breast. 3D information of the surface of the breast can be acquired using active or passive methods. The first act by projecting energy onto the breast and use its reflection to retrieve the surface information, while in the second 3D reconstructions of the breast are obtained by combining information of multiple images acquired with simple cameras. Among active methods, high resolution systems such as the 3dMD [14], which acquires multiple high-resolution images from several angles simultaneously, or other 3D laser scanning systems have been used to model breast surface. But these systems have an inherent disadvantage in the clinical set: they are expensive, large or have to be fixed in one place. They require a dedicated space to be properly operated and non-clinical specialized staff. As consequence, low-cost and easy to use alternatives have been explored, with examples using either more affordable active methods such as RGB-D cameras [8], or passive methods such as stereoscopy [6]. Breast surface representations can also be obtained from radar-based systems, in which the attenuation of a wideband pulse transmitted towards the breast is received by anthems. The received signals consist of two major contributions: the signal attributed to the skin reflection, and the signal from the internal structure of the breast. The skin reflection signal is used for the 3D breast surface reconstruction [12].

Regardless of the selected method to retrieve the surface information of the breast, registration algorithms have to be used to generate 3D representations of the breast surface. For instance, one of the challenges of using RGB-D cameras is the conversion of depth-map information to close meshes. In these strategies, breasts are usually imaged with the patient in a standing position, which frequently results in the lack of depth information for infra-mammary regions, particularly in the case of breasts with large curvatures. On the other hand, for stereoscopy-based methods, the challenge is to find corresponding points in multiple views images: occlusions may appear, and the lack of texture and salient features on the surface of the skin hardens the mapping process. This shows that despite the numerous available registration techniques, there is still room for improvement. Moreover, 3D breast reconstructed data are usually 3D point clouds, which difficults subsequent attempts of mapping surface information with breast interior data extracted from image radiological registration.

5 Conclusions and Future Work

Advances in imaging techniques have resulted in an increasing number of breast images. However, these images are often acquired with different views, modalities or at different times, which makes image registration an important step to convey multiple and complementary information into a single coordinate system to ease the understand and analysis of data. Yet, literature review evidences a gap for 3D breast image registration methods that combine interior and surface data of breast, and the use of 3D point clouds representing the breast surface can also pose extra registration challenges. Besides, the mapping of interior and surface information is a fundamental step in the creation of patient-specific complete models of the breast, which can be used for planning breast cancer surgeries and predict breast deformations arising from breast cancer treatments. Therefore, future research in the area of matching 3D interior and surface breast models would not only result in contributions in the field of computer vision, namely with new methodologies for 3D registration from multimodal images, but would also allow researchers to move ahead in the creation of 3D planning tools for visualization, to better understand the effect of surgical removal of cancerous tissue. Such tool would help the communication between the physician and the patient, ultimately empowering patients to take an active role in a shared decision making process.

6 Acknowledgements

This work was funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] Eric Bardinet, Laurent D Cohen, and Nicholas Ayache. Superquadrics and free-form deformations: A global model to fit and track 3d medical data. In *Computer Vision, Virtual Reality and Robotics in Medicine*, pages 319–326. Springer, 1995.

[2] Yujun Guo, Jasjit Suri, and Radhika Sivaramakrishna. Image registration for breast imaging: a review. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 3379–3382. IEEE, 2006.

[3] Jaime S. Cardoso Hooshiar Zolfagharnasab and Hélder P. Oliveira. A 3d parametric model for breast data. In *21st Portuguese Conference on Pattern Recognition*, pages 40–41, 2015.

[4] Torsten Hopp, Matthias Dietzel, Pascal A Baltzer, P Kreisel, Werner A Kaiser, Hartmut Gemmeke, and Nicole V Ruiter. Automatic multimodal 2d/3d breast image registration using biomechanical fem models and intensity-based optimization. *Medical image analysis*, 17(2):209–218, 2013.

[5] Angela WC Lee, Vijayaraghavan Rajagopal, Thiranja P Babarenda Gamage, Anthony J Doyle, Poul MF Nielsen, and Martyn P Nash. Breast lesion co-localisation between x-ray and mr images using finite element modelling. *Medical image analysis*, 17(8):1256–1264, 2013.

[6] Nicole Lepoutre, Marlène Gilles, Rémi Salmon, Christophe Collet, Barbara Bass, and Marc Garbey. A robust method and affordable system for the 3d-surface reconstruction of patient torso to evaluate cosmetic outcome after breast conservative therapy. *Journal of Computational Surgery*, 1(1):1, 2014.

[7] Thomy Mertzanidou, John Hipwell, Stian Johnsen, Lianghao Han, Bjoern Eiben, Zeike Taylor, Sebastien Ourselin, Henkjan Huisman, Ritse Mann, Ulrich Bick, et al. Mri to x-ray mammography intensity-based registration with simultaneous optimisation of pose and biomechanical transformation parameters. *Medical image analysis*, 18(4):674–683, 2014.

[8] Hélder P Oliveira, Jaime S Cardoso, André T Magalhães, and Maria J Cardoso. A 3d low-cost solution for the aesthetic evaluation of breast cancer conservative treatment. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 2(2):90–106, 2014.

[9] Diogo Pernes, Jaime S Cardoso, and Hélder P Oliveira. Fitting of superquadrics for breast modelling by geometric distance minimization. In *Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on*, pages 293–296. IEEE, 2014.

[10] Hayley M Reynolds, Jaykumar Puthran, Anthony Doyle, Wayne Jones, Poul MF Nielsen, Martyn P Nash, and Vijay Rajagopal. Mapping breast cancer between clinical x-ray and mr images. In *Computational Biomechanics for Medicine*, pages 81–90. Springer, 2011.

[11] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[12] M Sarafianou, DR Gibbins, and IJ Craddock. A novel 3-d breast surface reconstruction algorithm for a multi-static radar-based breast imaging system. In *2011 Loughborough Antennas & Propagation Conference*, 2011.

[13] Demetris Stavrou, Oren Weissman, Anna Polyniki, Neofytos Papageorgiou, Joseph Haik, Nimrod Farber, and Eyal Winkler. Quality of life after breast cancer surgery with or without reconstruction. *ePlasty: Open Access Journal of Plastic Surgery*, 9, 2009.

[14] Chieh-Han John Tzou, Nicole M Artner, Igor Pona, Alina Hold, Eva Placheta, Walter G Kropatsch, and Manfred Frey. Comparison of three-dimensional surface-imaging systems. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 67(4):489–497, 2014.

Compression Methods on Emotion Identification - Preliminary Study

Susana Brás[1]
susana.bras@ua.pt
Jacqueline Ferreira[2]
jacquelineferreira@ua.pt
Sandra C. Soares[2]
sandra.soares@ua.pt
Armando J. Pinho[1]
ap@ua.pt

[1] IEETA, DETI
Universidade de Aveiro
Aveiro, Portugal

[2] Department of Education and Psychology
Universidade de Aveiro
Aveiro, Portugal

Abstract

Emotions are under all our decisions. Therefore, in order to adapt systems and environments, it is important to study methods and measures for automatic identification of emotions. Since the electrocardiogram (ECG) is related with the nervous system, it contains information about our emotional state. Emotions are highly subjective. So, in this work, we present a method for similarity evaluation between ECG records, with the goal to identify the emotion. To accomplish this goal, we implement a measure based on the notion of joint compression of two objects. It is expected that two objects from the same source (in this case emotion) will use more information from one to describe the other, resulting in a smaller value of the dissimilarity measure. Using this method, we obtained distinct values when we compare values from the same emotion or from different emotions, indicating that this is a viable measure to explore in order to build an automatic emotion identification system.

1 Introduction

Emotions are the basis of our decisions, they define ourselves, and our choices, so by emotional identification we may develop systems and platforms able to adapt to each person, minimizing the interaction, and maximizing the benefit of the environment.
An increasing interest is verified on ECG for both biometric and emotion identification.
The electrocardiogram (ECG) is an electrical signal that contains information about our heart activity. Due to the physiological response of our body to events, our signals change accordingly. Therefore, due to the circadian cycle, or some particular circumstances (*e.g.*, stress, fatigue, emotional state), alterations are present in rhythm and/or amplitude [8] of the ECG signal.

The ECG is "full" of changes, some due to fluctuations or noise [1, 8], which sometimes compromise the identification process. Therefore, an ECG biometric identification system was proposed [3, 4] based on compression methods (using the Kolmogorov complexity). In those works, we assumed that a method based on a parameter free data mining in conjunction with a non-fiducial method [5, 8] (without ECG delineation) will improve the results, because there will be a reduction on the pre-processing error.
The physiologic response of emotions depends on the emotion felt, and also includes modifications in different systems and organs [2]. The description, definition and sensation of emotion is subjective - it depends on each person and, therefore, our hypotheses is based on the emotion identification dependent on the person. Therefore, as first step, we identify the person making use of the method described in [4], and then we identify the emotion felt.

2 Methods

Ten individuals (all females, age range 20-28 years) participated in this study. The participants had no reported diseases, were not taking medication, and had no previous history of psychiatric or psychological disorders. Participants were recruited at the University of Aveiro and received course credits for their participation. They gave their written consent and were informed about the possibility of withdrawing from the experiment at any time. The study followed the guidelines of the Declaration of Helsinki and standards of the American Psychological Association.

In a within experimental design, participants were shown three types of video - disgust, fear and neutral (one each week). The disgust film contained disgust scenes (*Pink Flamingos*), the fear film contained horror scenes (*The Shining*), and the neutral film displayed a documentary about the Solar eclipse *(Easter Island - Solar Eclipse)*. The disgust and fear films had been successfully used to induce disgust and fear, respectively, in previous studies (e.g., [6], [11]). The duration of each film was 25 minutes and the order of presentation of the movies was counterbalanced. The baseline data was collected by the presentation of a 4 minutes film of a beach sunset with acoustic guitar soundtrack. Also, they were instructed to avoid looking away or shut their eyes if they found the films too distressing.
During the experiment all participants' cardiac function was monitored by the use of a MP100 system and the software AcqKnowledge (Biopac Systems, Inc.), sampling the ECG at 1000 Hz. The adhesive disposable Ag/AgCl-electrodes were fixed in the right hand, as well as in the right and left foot.
For method development, and following the previous presented method [4], the ECG signal was decimated to 500 Hz, using an eighth-order low pass Chebyshev Type I filter with a 200 Hz cutoff frequency.

2.1 Compression Method

The ECG is a real-valued numerical signal, which should be converted in a sequence of symbols for a correct application of compression methods. To accomplish this step, we used the SAX representation [7]. Basically, 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 representation. The method optimization is dependent on the alphabet size and on the w dimension of the series segments.
The implementation idea of compression methods in this context is to find a measure of dissimilarity between records [5]. The data representation was performed by the use of finite-context models [10]. Based on the concept of algorithmic entropy, these models are used to calculate a dissimilarity measure [9].
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 [9, 10]).
After person identification, the emotion template was designed based on the first 20 ECG segments from the emotional record of each participant. The emotion identification method performance was evaluated in all dataset, the training being the first 20 segments and the testing the last 4 segments. The emotion identification was based on the lowest NRC value.
The results were presented as the emotion recognition rate, calculated as

$$\mathrm{RR} = \frac{TP}{N}, \tag{2}$$

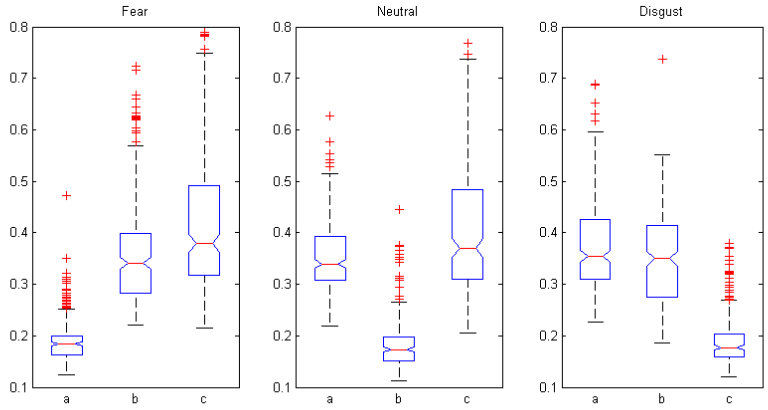


Figure 1: Boxplot representing the NRC measure in the comparison of each emotion (fear, neutral and disgust) with each other. The *a* box represents the comparison with fear, the *b* represents the comparison with neutral, and the *c* represents the comparison with disgust.

Table 1: Summary of the results obtained using the NCR measure in emotion differentiation. The results are presented as $mean \pm std$ of the calculated measure.

Emotion	Fear	Neutral	Disgust
Fear	0.189 ± 0.044	0.352 ± 0.074	0.412 ± 0.181
Neutral	0.357 ± 0.104	0.184 ± 0.053	0.387 ± 0.170
Disgust	0.464 ± 0.222	0.431 ± 0.187	0.189 ± 0.049

where TP is the correctly classified records and N is the total number of evaluated records.

3 Results

Following the NRC measure and the definition presented on [3, 4], in this work, it was decided to perform the method implementation with the same parameters. For quantization, an alphabet size of 20 and a window size of 3 points were chosen. Considering the finite context models, we assumed a γ of 0.2 and a combination of (α, k) pairs taken from set $\{(1/1000, 5), (1/1000, 6), (1/1000, 7), (1/1000, 8), (1/1000, 9)\}$ [3, 4].
Emotions are felt differently between persons. Therefore, a method to classify an emotion independently of the person, in theory, will present higher error. Therefore, in this preliminary approach we intended to present a similarity measure over emotions evaluating segments of ECG, considering that we are evaluating the same person.
This work idea is to present the pattern of the comparison between three emotions, when we compare emotion fear with emotion fear we expect a different NRC value than when we compare fear with emotion disgust or even fear with emotion neutral. So, in Figure 1, this difference in NRC values is evidenced by the presentation of a boxplot evaluating the comparison of each emotion with the three tested cases: fear (a), neutral (b), disgust (c). The statistics are presented in Table 1, evidencing that, when we compare different emotions, the NRC value increases (as already expected). These results demonstrate that the NRC may be used in the context of emotions. Also, because we need to previously identify the person, and following the previous results [3, 4], the same method may be used for both tasks, only with difference on the trainning template.
The emotion evaluation and automatic identification needs a further validation in a larger dataset, with heterogeneous gender. Also, positive emotions should be tested in this metric evaluation.

4 Acknowledgment

This work was supported by the European Regional Development Fund (FEDER) and FSE through the COMPETE programme and by the Portuguese Government through FCT - Foundation for Science and Technology, in the scope of the projects UID/CEC/00127/2013 (IEETA/UA), and CMUP-ERI/FIA/ 0031/2013, PTDC/EEI-SII/6608/2014. S. Brás acknowledges the Postdoc Grant from FCT, ref. SFRH/BPD/92342/2013. J. Ferreira acknowledges the Doctoral Grant from FCT, ref. SFRH/BD/ 85376/2012.

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] Foteini Agrafioti, Dimitrios Hatzinakos, and Adam K Anderson. ECG pattern analysis for emotion detection. *Affective Computing, IEEE Transactions on*, 3(1):102–115, 2012. ISSN 1949-3045.

[3] Susana Bras and Armando J Pinho. ECG biometric identification: A compression based approach. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 5838–5841. IEEE, 2015.

[4] Susana Bras and Armando J Pinho. Normalized relative compression (NRC) in ECG biometric identification. In *21st Portuguese Conference on Pattern Recognition (RecPad)*, Faro, Portugal, 2015.

[5] 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.

[6] Jasper H B de Groot, Monique A M Smeets, Annemarie Kaldewaij, Maarten J A Duijndam, and Gün R Semin. Chemosignals communicate human emotions. *Psychological science*, 23(11):1417–1424, 2012. ISSN 0956-7976.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] E P M Vianna and D Tranel. Gastric myoelectrical activity as an index of emotional arousal. *International Journal of Psychophysiology*, 61(1):70–76, 2006. ISSN 0167-8760.

A simple Net for a Deep Problem - Emotion Recognition

Ana Laranjeira[1]
afolgado@student.dei.uc.pt
Xavier Frazão[1]
xavier.frazao@eyeseesolutions.com
André Pimentel[2]
andre.pimentel@eyesee.pt
Bernardete Ribeiro[2]
bribeiro@dei.uc.pt

[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

There are few emotions that can be translated into facial expressions. These so called basic-expressions: *Anger, Disgust, Fear, Happiness, Sadness, Surprise*, according to Ekman's studies proved to be consensual within any culture. This fact allowed them to be present in the majority of expressions recognition systems, including the present one. In order to assemble a system capable of recognizing these expressions, we present a robust way of tackling the learning feature process, by introducing an improved version of the classic Convolutional Neural Network (CNN) - LeNet-5. Here we proved that net simplicity was better suited for the system constraints (dataset dimension, faces size and composition), comparing its performance with deeper networks (GoogleNet, AlexNet).
Assuming LeNet-5 as the baseline, we performed more experiments by refining their *hyperparameters* and focusing on three types of weight fillers *gaussian, unitball, xavier*. We used the Cohn-Kanade Extended (CKP) dataset for testing our proposed CNN model along with an augmented version due to their demonstrated effectiveness in the seven basic expressions. Moreover, we built a real-time video framework using our model (a version of LeNet-5) to reinforce the idea of robustness related to simple networks and the results are promising.

1 Introduction

The information that can be retrieved from emotion detection and recognition technology can increase the market competitiveness in a wide range of applications, specially in the marketing industry which centers its activity on digital social interactions between branding and the end-consumer. Therefore, there is an invested interest in Automatic Facial Expressions Recognition (AFER) systems. In the past few years some systems have revisited a concept - Convolutional Neural Networks (CNNs) in order to overcome most of the drawbacks related with the input. These network schemes make the systems more robust in the presence of variance and more scalable to real-time issues. The annually contest- Emotion Recognition in the Wild Challenge (EmotiW), is defining the state of the art in deep AFER systems. From the EmotiW challenge (2015), we can highlight two versions with expressions as subject: one uses static images from Static Facial Expressions in the Wild (SFEW); and the other uses an acted point of view resorting to an Acted Facial Expressions in the Wild (AFEW) dataset. Among the static-image approaches, a particular project [6] proposes a 3-way detection of the face with a hierarchical selection from the Join Cascade Detection and Alignment (JDA), Deep CNN-Based (DCNN) and MoT along with a simple network (11 layers). These detectors are processed in a multiple network framework in order to enhance the performance. It also includes a pre-processing phase to improve accuracy, which might be considered a drawback in the classification response. When considering video as resource there is an interesting approach [1] which introduces a Synchrony Autoencoder (SAE) to overcome spacio-temporal issues, by extracting local image features together with an hybrid network - CNN-RNN.

2 Dataset Augmentation

The quality of an AFER system relies significantly on dataset choices, therefore we used a standard in AFER systems and a reliable source - **Cohn-Kanade Extended** dataset [4] (CKP). The CKP is labeled between $0-7$ corresponding to *Neutral, Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise* and contains 593 sequences across 123 subjects with posed and non-posed (candid) expressions captured into a $640x \times 490$px or 640×480px frame, depending on the channel. We extended the prototypic expressions by including the class *Contempt*, mainly because it was reported to be found above 75% both in Western and non-Western cultures [2]. On the other hand, the neutral face was not considered in the training stage since it is hardly present in video-based classification. Only images with labels and in the peak of expression (apex state) were considered (1631 images). The CKP dataset was split into 70% for training; the remaining was taken for the validation and test phases. In order to feed properly the network, the **CKP** set of images were **augmented** with random perturbations, based on the expressive results [6]. The perturbation set, skew, translation, scale, rotation and horizontal flip worked separately in order to achieve a wider set, instead of the proposed overlapping method. Skew parameters were randomly selected from $\{-0.1, 0, 0.1\}$, translation parameters were sampled from $\{0, \delta\}$, where δ is a random sample from a $[0,4]$ set, scaling uses a δ value to define a random parameter $c = 47/(47-\delta)$ and the rotation is dependent on the angle sampled randomly from $\{-\pi/18, 0, \pi/18\}$. The final augmentation version has 978, 288 and 132 images per class (of seven) in training, validation and test phases, respectively. We included a set of images to the test phase, populated with frames from a real-time video framework composed by a Viola-Jones *OpenCV* [1] face tracker [5] along with our classifier. Images were resized to the classifier input shape (224×224px), captured within 35 frames per second, and classified in 0.250s (average including cropping process) into an expression displayed in the command line.

3 Simple Net - our version of LeNet-5

Our proposed model was developed using LeNet-5 as baseline. Its current stage is depicted in Fig. 1. Fixing the baseline network involved some preliminary experiments. In the initial developed phase we tested three prominent networks from the state of the art: the AlexNet from classifications of the ILSVRC2012 challenge, the recent GoogLeNet and LeNet-5. We used the non-augmented dataset which is small enough for a CNN input and therefore a candidate to make a sanity check on the *hyperparameters*. In this context, GoogleNet and LeNet-5 both passed the test, overfitting with an high accuracy between [0.9,0.92] in training, whereas the validation loss computed by summing the total weighted loss over the network (for the outputs with non-zero loss), reached a value of 0.8. The preliminary classification experiments ran over a short amount of training time and the best gains in the test set came from the LeNet-5 as depicted in Fig. 2. Since the results over the training and validation set had a retarded loss decay, no further experiments on GoogleNet and AlexNet were developed because training memory and processing time would become an issue. Considering these marks, a new version of LeNet-5 was explored. Moreover, LeNet-5 baseline parameters were addressed from the *Caffe* [2] standards. The classic LeNet-5 architecture [3], is a combination of seven layers represented by a convolutional layer with 16 feature maps with a 5×5 kernel size, followed by a sub-sampling by half. This sub-sampling (pooling layer) and the next convolutional layer (16 feature maps) are not connected in order to break symmetry. The network includes also another pooling stage, maintaining the 16 feature maps and the kernel to 5×5.

[1] http://opencv.org/
[2] http://caffe.berkeleyvision.org/

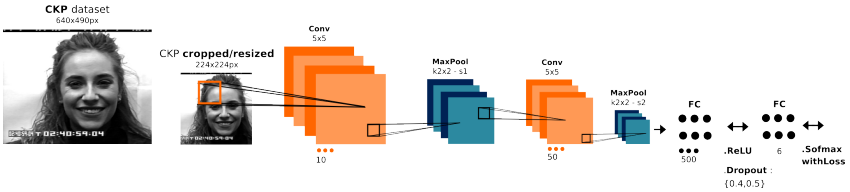


Figure 1: Flowchart of the different stages of our CNN model, Our version of the classic LeNet-5 - Lenet Ov

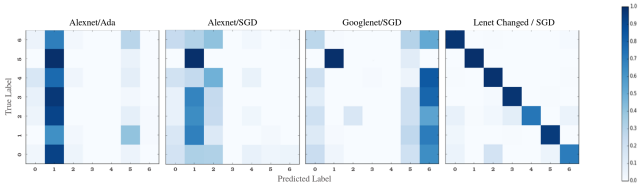


Figure 2: Normalized Confusion Matrices - (a) Alexnet/Ada; (b) Alexnet/SGD; (c) Googlenet/SGD; (d) Lenet changed/SGD.

It includes another convolutional full sized (kernel 1×1), mapping 120 units before the last full connected layers with 84 and 10 neurons, respectively (since LeNet is addressed to digit recognition).

Our model is composed by an initial convolutional set with a 5×5 kernel size and 20 feature maps plus a shared bias ending up with 520 parameters. The next layer or Pooling Layer performs a downsampling with a maximum value of a 2×2 kernel size. This process is repeated except for the pooling stride which changed to 2 and the convolution process expecting 50 instead of 20 feature maps, augmenting the parameters to 1300. To reduce the training error rate, the full connected layer that follows is connected with an *Rectified Linear Unit* (ReLU), containing 500 filter numbers. We introduced a *Dropout* between the full connected layers, discarding some units or neurons (between 0.4 and 0.5 percent of their connectivity) in the forward pass. The main goal of this procedure is to manage the overfitting, preventing co-adaptation. Finally, the last full connected layer is responsible for shrinking the feature maps to our class problem - 7 (expressions). The weights presented to the net follow the *xavier* type except for the first convolutional layer which is set between *gaussian, unitball, xavier*. The *gaussian* filler only chooses values according a gaussian distribution, limiting non-zero inputs up to 3, and the standard deviation assume a 0.01 value (increased from the default 0.005). The *positive unitball* fills a blob with values between $[0,1]$ such that $\forall i \sum_j x_{ij} = 1$. Finally, the *xavier* type (weight filler) initializes the incoming matrix with values from an uniform distribution within $[-\sqrt{\frac{3}{n}}, \sqrt{\frac{3}{n}}]$, where the n is the number of the input neurons. Our version of the network is optimized with a stochastic gradient descent solver, since the AlexNet training with Adagrad (see Fig. 2) was not expressive. Our parameters fit the hyperspace, with 0.09 for the momentum, 0.0005 of weight decay and a initial learning rate of 0.001, dropping a factor of 10 in the last 10% of iterations. To evaluate properly the robustness of the system, we scaled to a sequence-based model by introducing a video framework in testing phase. Despite this framework not being tested with 132 different subjects as it should be in an ideal version (equal to the number used to test CKP dataset), this simple test is enough to infer the learning stage.

3.1 Evaluation

From the three versions tested, the best results were obtained using the *UnitBall* configuration trained over 100000 iterations. It achieved a top recognition with a F1-score of 0.906 ($\sigma = 0.067$), according to an average based on values from Table 1. The *UnitBall* version of the model also achieved a 90% accuracy on the test set which is close to current state of the art (with 93%) accuracy using CKP dataset. Our final tests involving a framework built with real-time sequence-based images revealed that *Contempt* is the expression captured more frequently, even when it is not meant to happen. Moreover, the expressions of Anger, Happiness and Sadness were not properly learned. These results may occur due to the method used to capture the images, namely the Viola-Jones

	Gaussian (a)			**UnitBall** (b)			Xavier (c)		
	Prec.	Rec.	F1	**Prec.**	**Rec.**	**F1**	Prec.	Rec.	F1
Anger	0.747	0.985	0.850	0.949	0.985	0.967	0.833	0.985	0.903
Contempt	0.949	1	0.974	1	1	1	1	1	1
Disgust	0.652	0.894	0.754	0.820	1	0.901	0.830	1	0.907
Fear	0.820	0.864	0.41	0.924	0.917	0.920	0.897	0.924	0.910
Happiness	1	0.591	0.743	0.972	0.780	0.866	0.943	0.750	0.835
Sadness	0.840	0.795	0.817	0.881	0.955	0.916	0.884	0.924	0.904
Surprise	0.716	0.477	0.573	0.861	0.705	0.775	0.949	0.705	0.809

Table 1: Precision, Recall and F1 from three above versions (a)(b)(c).

algorithm may present some drawbacks in the presence of motion. Due to the fact that only static images were our goal, no further improvements were made in this section. Despite, the results not matching the desirable performance of the static test version and the difficulty in recognizing "emotions in the wild", our model proved that the process of learning with static images is scalable to sequence-based schemes since some of the expressions were recognized (such as fear and surprise). On the other hand, we encountered difficulties also raised by the current state of the art, particularly with the misguidance of some classes (such as Disgust in the test set [6]).

4 Conclusion

After several enhancements, we ended with a simple and significant solution for the AFER system network. The results from the crafted network attained 90% of accuracy in the test set (with static images). Both the inclusion of a Dropout layer in architectural decisions and the augmentation performed on the dataset helped to achieve better performance. Although the results are preliminary, the deep inference model also proved some robustness when performing with a sequence of video frames, at least for some of the facial expressions. Finally, we have acknowledged that there is space in sequence-based systems for improvement and it might include new tracker techniques or network adjusts.

References

[1] S Ebrahimi, Michalski, and other. Recurrent Neural Networks for Emotion Recognition in Video. In *Proc. of the ACM on Int Conf on Multimodal Interaction*, pages 467–474. ACM, 2015.

[2] Paul Ekman and Karl G. Heider. The universality of a contempt expression: A replication. *Motivation and Emotion*, 12(3):303–308, 1988.

[3] Y. Lecun, Bottou, et al. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 86(11):2278–2324, 1998.

[4] P. Lucey, Cohn, et al. The Extended Cohn-Kanade Dataset (CK+):A complete dataset for action unit and emotion-specified expression. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 94–101, 2010.

[5] Paul Viola and Michael J. Jones. Robust Real-Time Face Detection. *Int. J. Comput. Vision*, 57(2):137–154, 2004.

[6] Z Yu and C Zhang. Image based Static Facial Expression Recognition with Multiple Deep Network Learning. In *Proc. of the ACM on Int Conf on Multimodal Interaction*, 2015.

Landmines detection using thermal infrared sensors

Jorge Leitão Pimenta
pimenta.jl@mail.exercito.pt
Departamento de Ciências e Tecnologias de Engenharia
Academia Militar – Lisboa, Portugal

José Silvestre Silva
jose.silva@academiamilitar.pt
Departamento de Ciências e Tecnologias de Engenharia
Academia Militar – Lisboa, Portugal

José Bioucas-Dias
bioucas@lx.it.pt
Departamento de Engenharia Electrotécnica e de Computadores
Instituto de Telecomunicações e Instituto Superior Técnico,

Abstract

This work explores the detection of landmines using thermal images acquired in a military scenario and proposes a methodology following two main phases: acquisition of thermal images and its processing. In the first phase, several experiences were prepared to analyze the factors that influence the quality of the detection. In the second phase the acquired images are classified using the K-Nearest Neighbours (KNN) and the Support Vector Machine (SVM) classifiers.

1 Introduction

This work explores the detection of landmines using thermography. The problem of landmine clearance is current, complex and demanding due to a multiplicity of factors to consider at the time of detection. Landmine detection is a delicate activity due to the constant danger of explosion, when specialized engineers proceed with its inactivation and removal. This problem results in human victims and time and money invested in demining.

In the last decade many types of technology have been studied in the area of the physics of sensors, signal processing and robotics for the detection of landmines.

Roughan [1] uses the fusion of two sets of images, each set corresponding to a different infrared spectral range. One range is located between 3-5 μm and the other one between 8-12 μm. The obtained images allowed the extraction of features based on grayscale statistics and rotationally invariant statistics. Five classification algorithms were applied, two based on a single sensor (one for each spectral range) and three based on fusion techniques.

According to Paik [2] the detection of landmines can be performed through volume effect and surface effect. Thermal images were obtained with a sensor that detects infrared radiation in the spectral range between 3-5 μm. The techniques used for the detection of landmines were: filtering, feature extraction, contrast enhancement and segmentation.

Padmavathi [3] used Gaussian and Sobel filters for contours detection and noise removal that allow contrast enhancement. The segmentation was done with the H-maxima transformation algorithm followed by the KNN classification algorithm considering one class for landmine presence and other for the background.

The methodology presented by Suganthi [4] is based in a retro propagation neural network. Initially a pre-processing is performed with histogram equalization and Wiener filter to improve the contrast and remove the noise. The classification is performed through an Artificial Neural Network (ANN) with retro-propagation using a multilayer perceptron (MLP) topology.

Lee [5] presents a similar processing technique to the one presented by Paik [2]. However, such processing was applied to four different case studies. The images were obtained with three different sensors: 3-5 μm, 8-12 μm and 8-9 μm and allowed the identification of all landmines installed, in an experiment that simulated a real situation, with irregularities in the soil and the presence of vegetation.

Gader [6] proposed and evaluated general methods for detecting landmine signatures in ground penetrating radar using hidden Markov models. The models were successfully tested at two different locations contained approximately 300 landmine signatures (plastic or non-metal).

The goal of this work is to develop and implement a methodology for the detection of landmines with two steps: acquisition of thermal images considering the factors that influence the quality of the detection, and locating using the K-Nearest Neighbours (KNN) and the Support Vector Machine (SVM) classifiers.

2 Methods

2.1 Minefield Creation

Four experiments in different conditions were created to study the factors that influence the detection: characteristics of the soil, characteristics of the buried object, position of the buried object and thermal energy [7-9].

The first experiment conducted in a room seeks to compare the influence of different types of soils and different materials in scenarios conditions where there is a lower thermal energy. Three boxes measuring 0.8×0.8×0.25m were made, simulating three different minefields, each with a different soil type. One box has black soil, another sandy soil and the last organic soil [7].

The second experiment was conducted in the outside of the building, in the Military Academy (Lisbon, Portugal). The three boxes used in experiment 1, as well as the same types of soil that were placed within each, were also used in this experiment.

The third experiment consisted in using two minefields, one with sparse vegetation, and other with high dense vegetation.

In experiment 4, four minefields measuring 5×5m with real mines were created. In each field five mines were installed, with one being placed in each corner and one in the centre of the minefield.

2.2 Mine detection

Detection of landmines through thermal infrared images is obtained through the differences in temperature and/or spectral colour between landmine pixels and those of the background.

Characteristics that are informative with respect to mine class were extracted from ROI (Region Of Interest) within a sliding window that scans all images. The obtained feature vectors were then classified with the K-Nearest Neighbours (KNN) and Support Vector Machine (SVM) [10-13] supervised classifiers.

3 Results and discussion

3.1 Minefield Creation

Four different experiments were conducted with the aim of studying factors that affect mine detection via thermal imaging.

Experiment 1 was conducted in a room with the objective of simulating an exterior situation where there was no direct exposure to sunlight. All of the images were obtained at a height of approximately 2 m, and the resolution of the images about 12 pixels per centimeter of terrain (see figure 1).

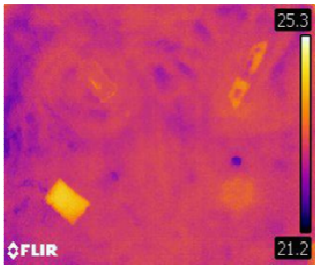


Figure 1: Box with organic soil in experiment 1, with objects on the surface (thermal imaging).

Experiment 2 is a replica of the previous experiment, except that it is conducted in the exterior. Figure 2 shows the thermal imaging for objects at a depth of 1.5 cm. At this depth just the box with sandy soil allows the regions where the objects are buried to be cleared identified.

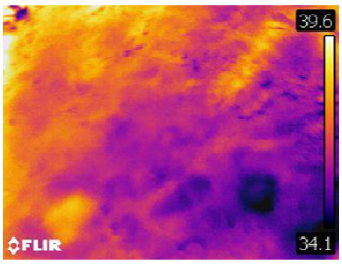


Figure 2: Box with sandy soil with objects at a depth of 1.5 cm (thermal imaging).

In experiment 3, five elements were placed within the field, four in each corner and one in the centre.

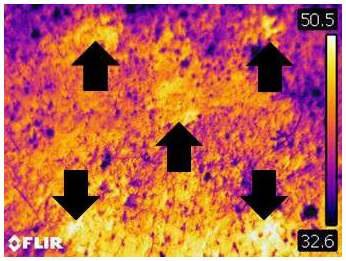


Figure 3: Field 1, thermal image after 11 days.

The thermal imaging of figure 3 shows that there are regions that have an apparent higher temperature, pointed to by the black arrows. All of the corresponding targets are visible in the thermal imaging, despite the edges not being clearly defined enough to allow an exact correspondence between the mine heat signatures and the background.

In the fourth experiment several minefields were made to be as close to reality as possible. Each field was 5×5m, and had five mines in each. The terrain used had irregularities, some vegetation and human intervention. Figure 4 presents the image of one of the minefields.

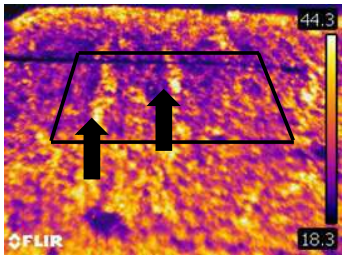


Figure 4: Infrared image of a minefield used in experiment four.

It is possible to visually identify in the previous minefield the mine in the centre and the mine in the lower left hand corner and which are indicated by the black arrows. The mine in the centre is on the surface and camouflaged by vegetation. The second mine is also at the surface, and it is primarily metallic.

On the other hand, there are zones without vegetation that appear to have high temperatures; this makes buried mine detection a hard task.

3.2 Procedure for mine detection

From the set of acquired images, with dimension of 225×300 pixels, the five more informative regarding the presence of mines, using a visual criterion, were selected for further processing.

Two classes were defined: mine and background. A total of 92 characteristics were calculated for each ROI using 8×8 pixels window in a total of 2292 ROI's for training and the remaining ROI's for testing.

To study the performance of the KNN classifier as a function of the parameter K, the Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) was applied. The classifier was implemented using 10-fold cross-validation.

The performance of mine detection using the KNN classifier and characteristics selected by SFS had an average accuracy of 85% for two characteristics, and 90% for 40 characteristics.

The performance of the SVM classifier for the different kernels was evaluated. The best result was for SVM using SFS, with linear kernel giving an accuracy of 87%, compared to using Gaussian kernel with 35 characteristics, which gave a maximum accuracy value of 85%.

4 Conclusions

Demining is a current, complex, and demanding problem because of the many factors that have to be taken into consideration during mine detection. As well as the human toll, time, money invested in demining, mines also restrict access to large areas of terrain that could be used by the local population. It is therefore crucial that new methods are found so as to solve this global problem, both in terms of sensors and processing algorithms.

In this work two areas related to mine detection were studied: thermal imaging obtained from experimental cases and minefields, and image detection based on classification. Four experiments were conducted, in which mine fields were set up to obtain a set of images that would enable the capacity to detect landmines.

From the results obtained, it was concluded that without external thermal stimulation, landmine detection is difficult as the corresponding signature is in most cases equal or lower than the background.

Sandy soil gave better results than either black or organic soil. This is essentially because the sandy soil is more homogenous which gave images with less distortion producing a better classification of minefield areas.

References

[1] D. Roughan. "A Comparison of methods of data fusion for land-mine detection", in *Int. Workshop on Image Analysis and Information Fusion*, 1997.

[2] J. Paik, C. Lee, M. Abidi, "Image processing-based mine detection techniques: a review," *An International Journal Subsurface Sensing Technologies and Applications*, 3: 153-202, 2002.

[3] G. Padmavathi, P. Subashini, M. Krishnaveni, "A generic framework for landmine detection using statistical classifier based on ir images," *International Journal on Computer Science and Engineering*, 3:254-262, 2011.

[4] G. Suganthi, D. Korah, "Discrimination of mine-like objects in infrared images using artificial neural network," *Indian Journal of Applied Research*, 4:206-208, 2014.

[5] C. Lee, "Mine detection techniques using multiple sensors." *The Project in Lieu of Thesis, Department of Electrical and Computer Engineering, University of Tennessee at Knoxville*, 4-18, 2000.

[6] P.D. Gader, M. Mystkowski, Yunxin Zhao. "Landmine detection with ground penetrating radar using hidden Markov models", IEEE Trans. on Geoscience and Remote Sensing 39(6): 1231-1244, 2001.

[7] R. Dam, B. Borchers, J.Hendrickx, et. al. "Controlled field experiments of wind effects on thermal signatures of buried and surface-laid land mines," in *Proceedings of SPIE, The International Society for Optical Engineering*, 541(5):648-657, 2004.

[8] C. Santulli, "IR thermography for the detection of buried objects: a short review," *Department of Electrical Engineering, Università di Roma – Italy*, 2007.

[9] R. Dam, B. Borchers, J. Hendrickx, "Strength of landmine signatures under different soil conditions: implications for sensor fusion," *International Journal of Systems Science*, 36:573-588, 2005.

[10] B. Pathak, D. Barooah, "Texture Analysis based on the gray-level co-occurence matrix considering possible orientations", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 2:4206-4212, 2013.

[11] P. Horton, K. Nakai. "Better prediction of protein cellular localization sites with the k nearest neighbors classifier", *Intelligent Systems for Molecular Biology*, 1:147-152, 1997.

[12] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors*, 12:12489-12505, 2012.

[13] T. N. Phyu, "Survey of classification techniques in data mining", in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 1:730, 2009.

Discriminative Directional Classifiers: Logistic Regression and K-Nearest Neighbors

Kelwin Fernandes
kafc@inesctec.pt
Jaime S. Cardoso
jsc@inesctec.pt

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

Abstract

In different areas of knowledge, phenomena are represented by directional -angular or periodic- data; from wind direction and geographical coordinates to time references like days of the week or months of the calendar. These values are usually represented in a linear scale, and restricted to a given range (e.g. $[0, 2\pi)$), hiding the real nature of this information. Therefore, dealing with directional data requires special methods. So far, the design of classifiers for periodic variables adopts a generative approach based on the usage of the von Mises distribution or variants. Since for non-periodic variables state of the art approaches are based on non-generative methods, it is pertinent to investigate the suitability of other approaches for periodic variables. We propose discriminative directional versions of the Logistic Regression and K-Nearest Neighbors models able to deal with angular data, which do not make any assumption on the data distribution.

1 Introduction

Several phenomena and concepts in real life applications are represented by angular data or, as is referred in the literature, directional data. Some examples of directional information are the wind direction as analyzed by meteorologists, magnetic fields in rocks studied by geologists, geographic coordinates, among others [1, 4]. Also, some entities are usually referenced in an angular manner; gynecologists denote the location to perform a biopsy, when performing a colposcopic screening, using the angle formed by the vertical axis of the cervix. Another example can be found in the area of computer vision, where color is often defined in cylindrical spaces like the Hue-Saturation-Value (HSV) color space. However, directional information is not constrained to scientific contexts; on a daily basis we naturally use angular variables. For example, time is usually represented by hours, days of the week, day of the month, season, etc. This reference system is cyclic by nature. Directional variables are usually encoded as a periodic value in a given range (e.g. $[0, 2\pi)$, $[0^\circ, 360^\circ)$). This work focuses merely in this representation of directionality, where an angular variable is a real-value number with periodicity defined by a range.

Working effectively with directional data requires dealing with techniques that are aware of the angular nature of the information [4]. For example, 0 and 2π are indeed the same angle and their average is not π but 0. In this sense, directional statistics concerns the problems derived from using traditional linear statistics with this type of data [4]. In order to formalize the definition of a directional function, consider the predicate *dir* defined in the Eq. (1), where $\mathbb{N}$ is the set of integers and $\mathbb{B} = \{\text{true}, \text{false}\}$.

$$\begin{aligned} &dir : \mathbb{N} \longrightarrow \mathbb{B} \\ &dir(i) = \text{true}, \quad \text{iff the } i\text{-th feature is directional} \end{aligned} \tag{1}$$

We will say that the function f, with domain in $\mathbb{R}^n$, is directional with period $\vec{P}$ (i.e. the feature in the position i has period $\vec{P}_i$), if and only if the Eq. (2) holds, where non-directional features are assumed to have infinite period (i.e. $\neg dir(i) \Rightarrow \vec{P}_i = \infty^+$).

$$f(\vec{\theta}) = f(\vec{\theta} + \vec{k} \circ \vec{P}), \qquad \vec{k} \in \mathbb{Z}^n \tag{2}$$

Here on, we will restrict the periodicity of the directional values to $P_i = 1$, without loss of generality.

Supervised learning can be understood as the process of learning a function f based on so-called training data that comprises examples of the input vectors and their corresponding target values. In this work, we are interested in the learning task known as classification, where the target can take a finite number of values. These values are usually denoted as classes or labels and the input vector defines a set of features that describe objects in the domain of the function. As the result of a supervised classification task, we obtain a classifier, which is used to assign a class to an object that has not been seen at the training stage. Traditional models that do not take into account directionality may suffer drop of generalization in areas near to the period of the function. Furthermore, the function may return different decisions for different $\Delta + \vec{k} \circ \vec{P}$, $\vec{k} \in \mathbb{Z}^n$, and a fixed $\Delta \in \mathbb{R}^n$, despite all of them semantically represent the same angle.

Previous attempts to design classifiers for periodic data adopted a generative approach based on the von Mises distribution or variants [3, 4, 5]. López et al. proposed a directional naïve Bayes formulation [4]. Their contribution involves using the von Mises and von Mises-Fisher distributions for the directional variables instead of the classic Gaussian distribution. The effectiveness of this method relies on the independence assumption of the features and the adequacy of the von Mises distribution to model the behavior of the directional features.

Since state of the art approaches are based on non-generative methods for non-periodic variables, in this work we propose two discriminant approaches to classify directional data. Our contributions stand as a directional-aware version of the Logistic Regression and a directional-aware version of the K-Nearest Neighbors classifier. The former is the discriminant counterpart of the naïve Bayes classifier, previously used to address this problem.

2 Directional Logistic Regression

Eq. (4) defines the Directional Logistic Regression (dLR) model. This model can be understood as a Logistic Regression with a mapping from the original angular space to a linear one. This mapping is learned simultaneously with the feature coefficients. Hereinafter, the two possible labels belong to $\{0, 1\}$, and n is the number of features.

$$\begin{aligned} f(\theta) &= \frac{1}{1 + e^{-h(\theta)}} \\ h(\theta) &= \omega_0 + \sum_{i=1}^{n} \omega_i g_i(\theta_i) \\ g_i(\theta_i) &= \begin{cases} \sin(2\pi(\theta_i + \varphi_i)) & ,\text{if } dir(i) \\ \theta_i & ,\text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

Given the properties of the sine function, the model holds the directional condition. We have analyzed the expressiveness of this model in [1]. In order to find the best model for a given training set we used gradient descent by considering the derivatives with respect to each parameter (ω_i and φ_i).

3 Directional K-Nearest Neighbors

Adapting traditional K-Nearest Neighbor classifier to include directional awareness can be done in a straightforward manner. For instance, it is enough to consider a directional-aware distance without modifying the underlying nearest neighbors algorithm. For example, the directional version of the Euclidean distance between any two points $\theta^{(1)}$ and $\theta^{(2)}$ is defined as follows:

$$\text{dir-}L_2(\theta^{(1)}, \theta^{(2)}) = \left(\sum_{i=1}^{n} d\left(\theta_i^{(1)} \% P,\ \theta_i^{(2)} \% P,\ dir(i)\right)^2 \right)^{\frac{1}{2}} \tag{4}$$

$$d(a, b, \text{isdir}) = \begin{cases} \min(|a-b|, P - |a-b|) & ,\text{if isdir} \\ |a-b| & ,\text{otherwise.} \end{cases} \tag{5}$$

Table 1: Average accuracy per model using 5-fold cross-validation.

Dataset	GNB	vMNB	LR	dLR	KNN	dKNN
Colposcopy	74.51 ± 7.24	70.52 ± 7.11	74.21 ± 6.71	$\mathbf{81.16} \pm 6.50$	81.05 ± 6.36	80.53 ± 6.64
Behavior	47.60 ± 9.20	49.53 ± 8.99	82.72 ± 3.64	$\mathbf{82.79} \pm 3.74$	80.56 ± 3.66	80.52 ± 3.70
Arrhythmia	67.06 ± 4.03	67.05 ± 4.07	78.31 ± 3.99	$\mathbf{78.38} \pm 4.04$	64.69 ± 4.18	65.94 ± 3.70
eBay	77.45 ± 3.37	83.88 ± 3.75	62.33 ± 4.42	$\mathbf{84.86} \pm 3.21$	77.96 ± 3.40	80.50 ± 3.15
Megaspores	76.72 ± 2.54	76.61 ± 2.71	62.50 ± 0.00	$\mathbf{76.78} \pm 2.58$	73.75 ± 3.05	73.75 ± 3.05
Characters	70.94 ± 2.62	73.40 ± 2.99	94.99 ± 1.59	$\mathbf{95.77} \pm 1.35$	86.97 ± 2.12	91.06 ± 1.89
OnlineNews	55.37 ± 2.12	55.29 ± 2.03	56.25 ± 2.94	$\mathbf{56.26} \pm 2.95$	50.52 ± 3.07	50.52 ± 3.07
Continents	94.66 ± 0.72	94.90 ± 1.08	94.79 ± 0.74	95.87 ± 0.72	97.91 ± 0.51	$\mathbf{97.93} \pm 0.51$
Wall	45.69 ± 2.01	51.07 ± 2.79	58.06 ± 1.39	66.53 ± 1.29	$\mathbf{86.51} \pm 0.97$	86.03 ± 0.96
Temperature0	68.56 ± 0.83	69.99 ± 1.80	59.15 ± 0.78	56.14 ± 0.92	$\mathbf{73.49} \pm 2.99$	$\mathbf{73.49} \pm 2.99$

4 Experiments

In this section we detail the experimental evaluation of the proposed directional Logistic Regression (dLR) classifier and its non-directional version Logistic Regression (LR) against their generative counterparts von Mises naïve Bayes and Gaussian naïve Bayes classifiers [4]. These methods can be summarized as follows:

1. GNB: Gaussian NB classifier that models continuous variables using Gaussian distributions.
2. vMNB: NB classifier that models linear variables using Gaussian distributions and directional variables using von Mises distributions.

Both Logistic Regression variants had an initial learning rate value (α) of 0.1 and a maximum number 10,000 iterations, but most datasets required much less iterations to converge. The model was initialized using small random values ($\omega_i \in [-0.05, 0.05]$ and $\varphi_i \in [-0.05, 0.05]$). The regularization constant λ was chosen using cross-validation in the range 10^{-2} and 10^{2}. The number of neighbors in the KNN and directional KNN (dKNN) models was varied between 1 and 10 using cross-validation.

4.1 Experiments

We validated the advantages of the proposed approach using ten real life datasets [2]. For this purpose, we compared the naïve Bayes variations, the classic Logistic Regression, KNN and the directional versions of the Logistic Regression and KNN classifier proposed in this work. Multiclass instances were handled using a one-versus-one approach for both versions of the Logistic Regression. All the experiments detailed below were executed with a stratified 5-fold cross-validation technique (by preserving the percentage of samples for each class) and results of 40 different runs were averaged. Results of these experiments are summarized in Table 1, exhibiting average accuracy and standard deviation for forty independent runs. The best model for each dataset is represented bold.

When comparing generative models, we obtained similar results to those obtained by López et al., namely vMNB achieves similar or better results than the GNB in most datasets [4]. The directional version of the Logistic Regression classifier reports a broad and significant advantage when compared with the non-directional approach. Moreover, dLR achieved better results than its non-directional and generative counterparts in almost all datasets. The main disadvantage of the proposed model, when compared with its generative counterpart, is the computational time required in the training stage. While naïve Bayes approaches require basic fitting of statistical distributions, dLR is learned by means of an iterative procedure, with asymptotic complexity $\mathcal{O}(I \times |S| \times N)$, where I is the maximum number of iterations, $|S|$ is the number of samples in the training set and N is the number of features. However, once trained, dLR is computationally competitive as it has linear complexity on the number of features – $\mathcal{O}(N)$.

Although the KNN approaches didn't obtained the best results in general, it can be observed that introducing directionality awareness to the KNN model improved its performance in most cases.

5 Conclusions

Different concepts in real life applications are represented by directional variables. These concepts are not restricted to the scientific domain, but can be easily found in daily routines, such as representing of time in a periodic repetitive calendar (e.g. hour, day of the week, month, etc). Traditional classifiers, which are unaware of the angular nature of these variables, might not properly model the data. Thereby, some directional classifiers have been proposed in the past, most of them using generative approaches and the directional von Mises distribution [4].

In this work, we proposed two directional adaptations to discriminative classifiers being able to receive mixed data (directional and linear). These classifiers add to the classic Logistic Regression (LR) and to the KNN classifier awareness about the angular nature of the data. As we demonstrated in the experimental assessment of the proposed models with real data, they can achieve competitive results when compared against their non-directional versions and against previous generative directional approaches.

Therefore, the directional Logistic Regression (dLR) and the directional KNN classifier offer promising results when dealing with directional data, and there is room for future improvement. In the near future, we intend to extend this work to Support Vector Machines (SVM), by redifining the concept of the functional margin in periodic spaces.

Acknowledgement

This work was funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF), and also by Fundação para a Ciência e a Tecnologia (FCT) within PhD grant number SFRH/BD/93012/2013.

References

[1] Kelwin Fernandes and Jaime S Cardoso. Discriminative directional classifiers. *Neurocomputing*, 2016.

[2] M. Lichman. UCI machine learning repository, 2013. URL `http://archive.ics.uci.edu/ml`.

[3] Pedro L López-Cruz, Concha Bielza, and Pedro Larrañaga. The von mises naive bayes classifier for angular data. In *Advances in Artificial Intelligence*, pages 145–154. Springer, 2011.

[4] Pedro L López-Cruz, Concha Bielza, and Pedro Larrañaga. Directional naive bayes classifiers. *Pattern Analysis and Applications*, pages 1–22, 2013.

[5] Richard S Zemel, Christopher KI Williams, and Michael C Mozer. Lending direction to neural networks. *Neural Networks*, 8(4):503–512, 1995.

Predicting Student Performance with Data from an Interactive Learning System

Ana Gonçalves[1]
argoncalves@ua.pt
Ana Tomé[2]
ana@ua.pt
Luís Descalço[1]
luisd@ua.pt

[1] Center for Research & Development in Mathematics and Applications
University of Aveiro
Aveiro, PT

[2] Institute of Electronics and Informatics Engineering of Aveiro
University of Aveiro
Aveiro, PT

Abstract

Nowadays Interactive Learning Systems have been developed to provide students with new forms of practicing concepts. In this work we propose to predict if the student fails or succeeds in the introductory mathematics course based on the information collected by an interactive learning platform. The predicting models are based on binary support vector machines (SVM). As some of the collected data sets are unbalanced the study was conducted with suitable strategies to train this binary classifier.

1 SIACUA application and collected data

SIACUA - Sistema Interativo de Aprendizagem por Computador, Universidade de Aveiro - is a web application designed to support autonomous study. For each subject is defined a concept map with questions associated to each concept. The system is supplied with parametrized questions from PmatE (`pmate.ua.pt`) and MEGUA (`cms.ua.pt/megua`) projects. It implements a user model based on Bayesian networks. The student chooses a subject and the system presents a sequence of related questions. After every student response the system provides a feedback. The student gets to know if the answer is correct, and in case it is not, he also gets the solution. In case of the MEGUA set the student also receives a detailled guidance of how to solve the problem. The student may also choose only to see the solution without answering the question. For a detailed SIACUA description please refer to Fonseca master thesis [3]. During a student session with SIACUA, the following information is stored: the student identification, the ID of each question and information related with student's question/answers. The latter comprises: the time-stamp for each question, the time elapsed between question presentation and student answer; elapsed time for solution visualization; the type of student answer (0- incorrect;1-correct, 2- solution visualization). This work proposes a machine learning based methodology to interpret the interaction of students with SIACUA. The goal of the work is to see if the behavior of the students with the system is related with the sucess of the student in the course. Therefore it is proposed a feature extraction block which describes the student in terms of platform's use. Afterwards a classifier is used to predict if the student approves or fails. Naturally this scenario leads to unbalanced data sets, as it is expected that the majority of students that were submitted to final evaluation succeed. Then, the SVM classifiers will be studied, however using methodologies to train with unbalanced data sets.

2 Data set

Only the students that were submitted to final evaluation were considered. In table 1 it is presented the number of students per group (that were approved or that failed) and discipline.

Discipline	Approved	Failed	Total
Cálculo 2	187	140	327
Cálculo 3	250	62	312

Table 1: Students by discipline and group.

Note that, for *Cálculo 2*, the cardinality of Approved and Failed sets are similar, while for Cálculo 3 the Approved subset is about 80% of students.

2.1 Feature extraction

Each student of the data set is represented as a vector $\mathbf{x}$ then forming data point in a space of dimension M. The vector has dimension $M = 42$ and each vector entry corresponds to one characteristic computed with the data collected by the platform. Examples of characteristics are the number of days the student interacts with the platform, the number of answers, the average time to answer questions and so on. Analysing the collected data by day perspective, it became possible to have, for instance, the number of answers a student gives during assessment and regular days.

3 Classification with SVM

Support Vector Machine (SVM) is a reliable two class classifier based on the decision equation

$$g(\mathbf{z}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{z}) + b \Rightarrow \begin{cases} g(\mathbf{z}) > 0, & y = 1 \\ g(\mathbf{z}) < 0, & y = -1 \end{cases} . \quad (1)$$

The parameters determined by classifier training are: α_i, the support vectors $\mathbf{x}_i$ and labels y_i and b. The support vectors are the elements of the training set that have associated nonzero α_i. $K(\mathbf{x}_i, \mathbf{z})$ is the kernel function. Two possible kernels were considered: the linear kernel

$$K(\mathbf{x}, \mathbf{z}) = \mathbf{x}^T \mathbf{z}, \quad (2)$$

which determines that the separation plane is a hyperplane; and the RBF kernel

$$K(\mathbf{x}, \mathbf{z}) = \Phi^T(\mathbf{x})\Phi(\mathbf{z}) = \exp(\frac{\|\mathbf{x}-\mathbf{z}\|^2}{-2\gamma}), \quad (3)$$

in which case the separation surface is not linear.

4 Linear SVM and unbalanced data

The hyperplane determined by using unbalanced training data sets is in most of the cases more close to the majority class [1, 5, 7]. Different approaches have been proposed to deal with unbalanced data sets. In the following subsections some are described: Different Error Costs (DEC), Sythetic Over-sampling + DEC and z-SVM.

4.1 DEC

The Different Error Costs (DEC) proposed by Veropoulos *et al.* [7] assigns different error costs to the misclassification of objects from each class. The linear SVM with smooth margin minimization problem is given by

$$\min \quad \left(\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i\right) \quad \text{with} \quad y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0, \quad (4)$$

where ξ_i are slack variables for the misclassified elements. The C parameter controls the tradeoff between the number of misclassified objects and the margin width and it is an user defined parameter.

Let C^- and C^+ be the error costs for misclassification of negative and positive class objects, respectively. The expression for SVM with smooth margin can be re-written as

$$\min \quad \left(\frac{1}{2}\mathbf{w}^T\mathbf{w} + C^+\sum_i \xi_i + C^-\sum_i \xi_i + b\right), \quad (5)$$

subject to the same conditions as in(4). After training using the dual form of the defined optimization problem the Lagrangians and support vectors are available. Note that in the case of linear kernel it is possible to compute the normal to the hyperplane $\mathbf{w}$ that characterizes the decision surface (e.g. the hyperplane can be computed using the support vectors and Lagrangian values [4]). While in the the case of RBF, the support vectors and Lagrangian values have to be stored to the test phase.

4.2 Synthetic Over-sampling + DEC

The experimental results presented by Akbani *et al*. [1] suggest that the application of synthetic over-sampling techniques before training DEC have better performance. Two techniques have been studied: Sythetic Minority Over-sampling Technique proposed by Chawla *et al*. [2], known as SMOTE and SMOTE SVM proposed by Nguyen *et al*. [6].

In SMOTE, the new instances are generated over the line segments that connect the nearest neighbors of the minority class. In SMOTE SVM, first an SVM classifier is trained on the original data set. Then, new instances are generated based on the minority class support vectors of the resulting model. This way, the new instances are generated along the decision boundary. Figure 1 illustrates the results obtained by both techniques.

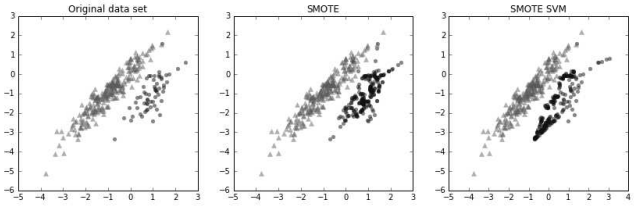


Figure 1: Synthetic over-sampling by SMOTE and SMOTE SVM.

Observing the middle graphic in figure 1 it can be seen that in the data set resulting of over-sampling with SMOTE, the minority class is denser having its instances uniformly distributed over the original instances area. While the data set resulting of over-sampling with SMOTE SVM (right graphic) has the new minority class instances concentrated along the decision boundary, where they are critical for estimating the optimal decision boundary.

4.3 z-SVM

Imam *et al*. [5] presented z-SVM. With this strategy, first an SVM classifier is trained. Then, the decision boundary is adjusted in order to correct the bias towards the majority class. Rewriting equation 1 by separating the support vectors from each class and then, magnifying the α values associated with the minority class support vectors by a small positive constant z, the z-SVM classifier is obtained:

$$g(\mathbf{z}) = z \sum_{i|_{y_i>0}} \alpha_i y_i K(\mathbf{x}_i,\mathbf{z}) + \sum_{i|_{y_i<0}} \alpha_i y_i K(\mathbf{x}_i,\mathbf{z}) + b \Rightarrow \begin{cases} g(\mathbf{z}) > 0, & y = 1 \\ g(\mathbf{z}) < 0, & y = -1 \end{cases} . \quad (6)$$

The z value is experimentally determined and is the one that maximizes the geometric mean g. Note that the minority class is considered as the positive one.

5 Results and Discussion

The simulations used the python software package scikit-learn[1] and the python library imbalanced-learn[2] for the synthetic over-sampling. Three data normalizations were considered: no normalization (none), min-max and z-score normalizations.

5.1 Evaluation

There are several performance measures. The usual one is accuracy, the proportion of objects correctly classified. Considering only the negative class object, the accuracy among these objects is the proportion of negative class objects correctly classified and designated specificity s. The proportion of positive class objects correctly classified is the sensitivity or recall r. For unbalanced data sets it is usual to use the geometric mean between specificity and recall as a performance measure

$$g = \sqrt{r \times s}. \quad (7)$$

5.2 Performance

Table 2 presents a summary of the results obtained for the Cálculo 3 data set.

	none		min-max		z-score	
Classifier	**acc.**	***g***	**acc.**	***g***	**acc.**	***g***
SVM linear	0.785	0.304	0.801	0	0.795	0.218
SVM RBF	0.801	0	0.801	0	0.801	0
DEC	**0.644**	**0.656**	**0.670**	**0.667**	0.587	0.655
SMOTE+DEC	0.603	0.644	0.628	0.651	0.487	0.579
S.SVM+DEC	0.641	0.643	0.715	0.602	**0.663**	**0.656**
z-SVM	0.385	0.422	0.401	0.344	0.494	0.344

Table 2: Accuracy and geometric mean g for unbalanced data.

It seems to be an advantage in the use of classifiers adapted for unbalanced data sets, with the exception of z-SVM. Recall that z-SVM's starting point is the decision boundary obtained by applying linear SVM. And, as shown by results in table 2, this model is much biased towards the majority class. So, it is not a surprise that the decision boundary adjustment provided by the z-SVM approach does not offer so good results.

DEC and SMOTE SVM + DEC classifiers are the ones with highest accuracy and geometric mean. So, for the studied data sets, the approaches for unbalanced data based on different error costs for missclassified objects present better results. For original data and min-max normalized data, the DEC classifier is the one with highest accuracy and geometric mean. For z-score normalized data, SMOTE SVM + DEC approach presents the best results. These results also suggest that is not always worth it the extra processing for data generation, before applying DEC classifier.

Although the Cálculo 2 data set is quite balanced, the approaches for unbalanced data sets were also tried. As expected there is no advantage in using these techniques on this set. The accuracy and g values are close to 0.72 for linear SVM and slightly better for RBF SVM. With unbalanced strategies, including the z-SVM, those values decrease (less than 3 %) . These results are according to what is reported in other works [1, 5].

6 Conclusion

Despite the application of strategies to deal with unbalanced data sets, the obtained results are fragile. It is believed that the amount of collected data is not enough to fully cover the diversity of the student behavior. So, concerning the work primary goal, it is not yet possible to predict about a student success. Hopefully, it becomes possible to determine a reliable prediction model as more data is collected.

The preliminaries results presented in this work show that the features are relevant for decision making and strategies for unbalanced data sets improve the classification of linear SVM.

References

[1] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. In *In Proceedings of the 15th European Conference on Machine Learning (ECML*, pages 39–50, 2004.

[2] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Int. Res.*, 16(1):321–357, 2002.

[1] http://scikit-learn.org

[2] http://github.com/scikit-learn-contrib/imbalanced-learn

[3] M.M. Fonseca. Modelo bayesiano do aluno no cálculo com várias variáveis. Master's thesis, Departamento de Matemática, Universidade de Aveiro, 2014.

[4] J. Han. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., 2005.

[5] T. Imam, K. M. Ting, and J. Kamruzzaman. z-SVM: An SVM for improved classification of imbalanced data. In *Proceedings of the 19th Australian Joint Conference on Artificial Intelligence: Advances in Artificial Intelligence*, pages 264–273, 2006.

[6] H.M. Nguyen, E.W. Cooper, and K. Kamei. Borderline oversampling for imbalanced data classification. *Int. J. Knowl. Eng. Soft Data Paradigm.*, 3(1):4–21, 2011.

[7] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of the International Joint Conference on AI*, pages 55–60, 1999.

Motion Recognition from Accelerometer, Gyroscope and ECG Data

Soraya Sinche
smaita@dei.uc.pt
Department of Electronic, Telecommunications and Networks
Escuela Politécnica Nacional

Bernardete Ribeiro
bribeiro@dei.uc.pt
Department of Informatics Engineering,
University of Coimbra

Jorge Sá Silva
sasilva@dei.uc.pt
Department of Informatics Engineering,
University of Coimbra

Abstract

In the last years, with the advances in smart personal things, make us to believe that the new generation of Internet of Things will become the human being as an integral part of the system. With this purpose the IoT technologies must be supported on Human-in-the-Loop Systems. In fact, the Human-in-the-Loop Cyber-Physical System (HiLCPS) considers human being as an integral part of the system. The data acquisition and sensing are fundamental parts within HiLCPS. In the data acquisition process, the principal devices that allow data collection are the sensors. There are several sensors that can be used to gather data on human activity and behavior such as the microphone, accelerometer, gyroscope, magnetometer, ECG, etc. With data collected is possible to infer various activities and moods of individuals and based on this information it is also possible to generate a feedback to improve quality life of a person. The reliability of data acquired is crucial, for this reason, the selection of a sensor or smart device is very important, and it must be taken depending on the type of applications to be implemented. Combining information from various sensors can represent an opportunity to improve the reliability and the accuracy of data. In this paper, we analyzed the accuracy of various classifiers for motion recognition. A public dataset that includes data from various sensors was analyzed. The data of accelerometer, gyroscope and electrocardiogram (ECG) were used in order to interfere human activity. The performance was evaluated applying the following classifiers: Decision Trees, Support Vector Machine (SVM), K-Nearest Neighbors (K-NN) and Hybrid Classifier. Finally, for the recognition of the four movement activities (no, slight, moderate and high movement), we obtained an accuracy of 99,40% with the K-NN classifier.

1 Introduction

Humans walk with many types of devices (e.g. smart-shirts, smartphones, smart glasses, smart watches, etc.). There are several sensors that can be used to gather data on human activity and behavior. These include the microphone, accelerometer, gyroscope, etc. These sensors allow one to conclude the activities and moods of individuals.

The public Mobile Health (MHealth) dataset [1][2] was used for the project, which contains information of sensors like the accelerometer, gyroscope, magnetometer and ECG. This dataset includes data collected from ten volunteers during several physical activities (standing still, sitting and relaxing, jogging, walking and running).

The information of sensors positioned on the subject's chest (accelerometer), on the right wrist (gyroscope) and two leads localized on the chest (ECG) were extracted from "MHealth" dataset. For analysis, the measures of 4 people were grouped. Each individual performed five activities: standing still, sitting and relaxing, jogging, walking and running. One of the activities was running by 1 minute with 50 Hz sampling frequency.

Four classifiers according to their parameters were analyzed with the values of database "MHealth" (accelerometer, gyroscope and ECG). These classifiers were: Decision Trees, SVM, K-NN [3] and hybrid classifier. The performance was evaluated with Weka toolkit [4] with high degree of accuracy.

Related work is presented in Section II. Section III shows the process of feature extraction and data analysis using different classifiers. The section IV presents the experimental results and their discussion. The section V includes the conclusions and future work in this research area.

2 Related Work

There are several activity recognition works. Ravi N. et. al. [5] presents the activity recognition as a classification problem. Some classifiers are compared using only a triaxial accelerometer sensor. The features are extracted individually for each axis (x, y and z). Meanwhile in [3], the features are based on the acceleration vector and integral (velocity) obtained from the accelerometer too. In [6], a novel motion recognition algorithm is presented, where the recognition movement is detected (e.g. arm moving up or down) using an accelerometer and a gyroscope. Nuguyen et. al. [7], propose the FE-AT (Feature-based and Attribute-based learning) approach. This proposal allows recognition of new activities using three public datasets (MHealth, DailyAndSport and RealDisp).

The present paper evaluates the contribution of different features in the motion level classification; the classification accuracy achieved with accelerometer, gyroscope and ECG features individually and also their combination.

3 Data Analysis for Motion Recognition

Activity recognition process is based on data analysis of sensors. This process interprets the sensor data to classify different activities [8]. The following considerations were made in the analysis: activities of standing still, sitting and relaxing were grouped such as motionless (NM), walking as a slight motion (SM), jogging as a moderate movement (MM) and running as a high movement (HM). The size of the data set contains a total of 15 370 values by person.

The features were extracted from the raw data of accelerometer, gyroscope and ECG with Matlab. In order to obtain reliable data, a windows size of 5.12 seconds (256 samples) was used with an overlapping of 50% in each case. Two features were calculated from each sensor: mean and standard deviation [5].

3.1 Preprocessing and Extraction of Features

Accelerometer data provide: time, acceleration along x axis (A_x), y axis (A_y) and z axis (A_z). Complementary to the three axes data, we can obtain the magnitude of the acceleration vector. The mean and standard deviation values were calculated for each axis (x, y and z) and acceleration module.

The gyroscope measures the rotation around one of the axes called angular rate (G_x, G_y, G_z), in degrees per second. The features calculated are the mean and standard deviation value for each axes.

ECG data are not of good quality to calculate specific parameters such as Heart Rate Variability; this is the reason why we only obtained the mean value and standard deviation of the full signal.

The data samples were labeled as "NM" no-movement, "SM" slight movement, "MM" moderate movement and "HM" high movement.

The process of standardization was used with the features vector. The size of features vector is equal to 475. Moreover, data were split into two sets: training (70%) and testing (30%) datasets.

3.2 Classification Models

After the data standardization process and with the use of the training data, four classifiers were analyzed using Weka: Decision Trees, SVM with Radial Basis Function as the Kernel function (SVM-RBF), K-NN and Hybrid Classifier.

For Decision Trees, J48 algorithm [9], pruned decision tree was set. In Weka, the value of confidence interval parameter (C) can be optimized with the CVParameterSelection.

In the case of SVM [6], the function of hyperplane can be linear, however in some cases it can to use a function more complex as Kernel function. The Sequential Minimal Optimization (SMO) was applied with RBF as the Kernel function, and the complexity parameter C as the gamma value (G) were optimized.

Lazy class with instance-based learning with parameter K (IBk) is used as K-NN. The parameter K specifies the number of Nearest

Neighbors (NN) used to the classifier in a test. The outcome is determined by majority vote. In this work, K value was optimized.

There exist some options that allow one to implement a hybrid classifier, for example: ensembles (Bagging and Boosting), Voting and Stacking [10]. The class vote was applied as meta-level classifier (Weka). Vote is a class for combining classifiers. We combined J48, SOM-RBF and K-NN in the hybrid classifier.

4 Experimental Results and Discussion

Firstly, the value of confidence interval parameter (C) for J48 classifier, the complexity parameter C, gamma value (G) for SVM-RBF and the K value with K-NN are optimized using CVParameterSelection and the matrix of features into Weka. Table 1 shows the ranges in which classifier are evaluated. All classifiers were run on data set in six different configurations:

(a) Attributes of Acceleration A_x, A_y and A_z (six attributes).
(b) Attributes of Acceleration vector (two attributes).
(c) Angular rate of Gyroscope G_x, G_y and G_z (six attributes).
(d) Attributes of two leads (ECG four attributes).
(e) Acceleration and Angular rate (A_x, A_y, A_z, G_x, G_y and G_z) with twelve attributes.
(f) Acceleration, Angular rate and ECG (A_x, A_y, A_z, G_x, G_y, G_z, $lead_1$ and $lead_2$) with fourteen attributes.

Classifier	J48	SVM-RBF		K-NN
Parameter	C	C	G	K
Range	0.1 – 0.5	2 – 8	0.01 – 0.1	1 – 16
Steps	5	4	10	4

TABLE 1. VALUES OF PARAMETERS ANALIZED

Table 2 presents the values optimized in each case with the six configurations. Then we present the results obtained using these values.

	Classifier	**J48**	**SVM-RBF**		**K-NN**
	Parameter	**C**	**C**	**G**	**K**
Configurations	(a)	0.1	4	0.01	1
	(b)	0.2	6	0.01	6
	(c)	0.1	8	0.01	1
	(d)	0.1	8	0.01	1
	(e)	0.1	8	0.01	1
	(f)	0.1	8	0.01	1

TABLE 2. VALUES OPTIMIZED

The average accuracy for four classifiers run with 10 fold cross validation is shown in Table 3[1]. In the case of hybrid classifier, the class Vote is used by combining J48, SVM-RBF and K-NN. Using the optimized values, all classifiers also were run for testing the six configurations a) - f), Table 4 shows these results.

Classifier	Accuracy (%)					
	(a)	(b)	(c)	(d)	(e)	(f)
J48	95.20	98.52	96.13	80.88	96.70	96.71
SVM	93.15	98.50	87.17	67.19	**98.80**	97.91
K-NN	**98.81**	**99.71**	**96.72**	**81.75**	**98.80**	**99.40**
Hybrid	97.60	**99.71**	96.41	80.89	**98.80**	98.81

TABLE 3. ACCURACY USING CROSS VALIDATION

In Table 3, the K-NN performs as the best configuration is visible. The second best is Hybrid Classifier. The best accuracy is obtained with module of accelerometer attributes. However, in Table 4 it is demonstrated when working with testing data, accuracy values in some cases stay far from the values obtained with cross-validation.

Classifier	Accuracy (%)					
	(a)	(b)	(c)	(d)	(e)	(f)
J48	82.86	98.57	22.14	41.43	79,29	79,23
SVM	**95.00**	**100**	**65.0**	43.57	88,57	91.43
K-NN	77.14	99.29	52.86	35.0	**92.14**	**92,86**
Hybrid	87.86	**100**	50.0	36.43	90.0	91.43

TABLE 4. ACCURACY USING TESTING DATA

[1] The standard deviations were omitted due to table format size in the paper.

5 Conclusions and Future Work

In this paper, an analysis of some classifiers for detection of movement was realized, using wearable sensors. To conclude, the sensor able to classify motion levels with better accuracy is the accelerometer.

From the results obtained, we can observe that using the attributes of the acceleration vector module, the best accuracy is achieved. Therefore, to implement mobile applications in real-time that need motion detection is advisable to work with the features of accelerometer.

Despite being important the information that allows inferring the level of movement of a person, there is a need to infer more complex situations. For example, for early warning systems in case of a physical attack or a theft, it is necessary to use several sensors capable of triggering alerts. So it is important to obtain information from various sensors, which could be integrated into equipment as a mobile phone. For this reason, we can not lose sight of the need to analyze parameters such as processing time and power consumption in the data processing phase. This allows us to select a classifier that can offer a good accuracy with a shortest processing time and low power consumption.

In this paper, the data processing was not performed in real time, but there are applications, where the data processing should be made in real-time. In these cases, an interesting extension of this work would be the analysis of parameters such as processing time and power consumption.

Acknowledgments

The work presented in this paper was partially financed by SENESCYT - Secretaría Nacional de Educacíon Superior, Ciencia, Tecnología e Innovacíon de Ecuador and the Escuela Politécnica Nacional de Ecuador.

References

[1] Banos, O., Garcia, R., Holgado, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., Villalonga, C. “mHealthDroid: a novel framework for agile development of mobile health applications.” Proceedings of the 6th International Work-conference on Ambient Assisted Living an Active Ageing (IWAAL 2014), Belfast, Northern Ireland, December 2-5, (2014).

[2] Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado, J. A., Lee, S., Pomares, H., Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. BioMedical Engineering OnLine, vol. 14, no. S2:S6, pp. 1-20 (2015).

[3] Casale P., Pujol O., and Radeva P., “Human Activity Recognition from Accelerometer Data Using a Wearable Device”, Springer-Verlag Berling Heidelberg (2011).

[4] Weka Toolkit (http://www.cs.waikato.ac.nz/ml/weka/)

[5] Ravi N., Dandekar N., Mysore P., Littman M., “Activity Recognition from Accelerometer Data”, American Association for Artificial Intelligence, p.p. 1541 – 1546 (2005).

[6] Varkey J., Pompili D., Walls T., “Human motion recognition using a wireless sensor-based wearable system”, Pers Ubiquit Comput,(2011).

[7] Nuguyen, L. T., Zeng, M., Tague, P., Zhang, J., “Recognizing New Activities with Limited Training Data”. In IEEE International Symposium on Wearable Computers (ISWC) (2015).

[8] Durmaz O., Ersoy C., “A Review and Taxonomy of Activity Recognition on Mobile Phones”, Bionanoscience, (June 2013).

[9] Ross Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA. (1993).

[10] Jain K., Duin Robert P.W., and Mao J., “Statistical Pattern Recognition: A Review”, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2000).

Multi-Object Tracking with Distributed Sensing

Ricardo Dias
ricardodias@ua.pt
Nuno Lau
nunolau@ua.pt
João Silva
joao.m.silva@ua.pt
Gi Hyun Lim
lim@ua.pt

IEETA
University of Aveiro
Aveiro, Portugal

Abstract

One of the main research goals on distributed autonomous agents in a Multi-Agent System is the development of mechanisms to form a better world model using information merging from different agents. In this paper, we present a solution for robust online and real-time multiple object tracking in a multi-agent system using information gathered by various agents over time, using COP-KMeans for clustering and Kalman Filtering for object state estimation.

The proposed solution was implemented on a real robotic soccer team and evaluated in the RoboCup Middle-Size League competitions. The robotic soccer was presented as one possible application for the ideas expressed on this paper.

1 Introduction

One of the main research goals on distributed autonomous agents in a Multi-Agent System (MAS) [5] is the development of mechanisms to form a better world model using information merging from different agents. It has already been demonstrated that distributed sensor fusion can enhance the belief by synergistically merging data from different agents [2] to derive a better approximation of the World Model than would be possible with each one individually [1].

When compared to centralised approaches, distributed systems present a major advantage: they are usually more resilient to failures. However, it is much more difficult to implement a real fully distributed system, when comparing with a centralised approach, in which one "master" agent takes control of a task that all the team will benefit from.

In this paper, we present a solution for robust real-time multiple object tracking in a multi-agent system using information gathered by various agents over time. For the purpose of this paper, we assume that each agent is able to detect object candidates and will focus on the integration of these observations into their world model.

The proposed solution was implemented on a real robotic soccer team and evaluated in the RoboCup Middle-Size League world championships. While having a huge potential for a variety of applications, MAS are extensively tested and benchmarked in RoboCup. The RoboCup Middle-Size League provides an excellent testbed for autonomous robotic teams in stochastic and highly dynamic environments. A soccer match cannot be overlooked as a testbed, since it resembles more the real world (complex and semi-unstructured) than a research lab. Although robotic soccer is presented as an example, the ideas expressed on this paper are not limited to this application area.

Following this Section, which introduces the problem and some concepts about the application, we will start by defining our implementation of the local object tracking in Section 2 that will define how tracking is done in each agent individually. In Section 3, we will present a methodology used to merge information from multiple agents to form a unified representation of the obstacles spread around the field. Then, we show some results and respective discussion in Section 4 and conclude the paper in Section 5.

1.1 RoboCup Middle-Size League

Among the RoboCup leagues, the Middle-Size League (MSL) is one of the most challenging in terms of rules and environment (Figure 1). In this league, robots play soccer autonomously in a 18x12m field with a standard size-5 FIFA ball. Each team can have up to 5 robots with maximum size of 50x50 cm base and 80 cm height and are not allowed to weight more than 40 kg. Currently, all teams participating in the league have an omni-directional drive system. The rules of the matches are based on the official FIFA rules, with a few required changes to adapt for the playing robots.

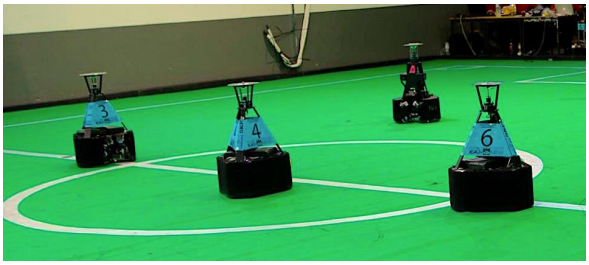


Figure 1: Middle-Size League final in RoboCup Portuguese Robotics Open at Bragança, Portugal

1.2 The CAMBADA Team

The developed work was accomplished within the MSL context, more specifically in the CAMBADA (Cooperative Autonomous Mobile roBots with Advanced Distributed Architecture) team [3], the MSL Robotic Soccer team from the University of Aveiro. This project started in 2003 and is currently coordinated by the IEETA IRIS group and involves people working on several areas from hardware (building the mechanical structure of the robot, its hardware architecture and controllers) to software (image analysis and processing, sensor and information fusion, reasoning and control).

2 Single-Agent Multi-Object Tracking

In a first instance, the integrator has to work with the local information it gets in the current cycle, only afterwards it is able to integrate information from other agents. In this Section, we show how we implemented our Object Tracking software module and discuss its usage in both obstacle tracking and ball tracking.

In this work we assumed that each track can be modelled as a Gaussian process in the 2D space, so each will contain a Kalman Filter with 4 state variables: x, y, $\dot{x}$, $\dot{y}$ - position and velocity in x and y axis, of which implementation is beyond the scope of this paper. Since robots are omni-directional, orientation was not considered.

2.1 Obstacle Tracking

Being soccer a very dynamic scenario, our robots need to be able to perceive the other robots around them, both opponents and their own team-mates. They are required to move around the field, either for repositioning or dribbling the ball, while avoiding contact with any other obstacle on the field.

The simplest approach would be a reactive one, in which the actions of the robot are defined by the obstacles seen at that agent cycle. However, because the robots move and rotate at very high speeds, there is always a percentage of false-positives and false-negatives, which justifies the need for an obstacle tracking method.

For each new agent cycle, the object tracker algorithm is updated with the observations on that cycle. This algorithm can be summarized in the following steps steps:

1. Predict new track positions
2. Assign observations to tracks
3. Update assigned tracks
4. Update unassigned tracks and purge old tracks
5. Create new tracks

2.2 Obstacle Sharing Criteria

Despite locally detecting and taking into account every observation in its decisions, each agent is very cautious about the information it shares. Since other agents on the team can use the shared information, it is very important that the information that an agent broadcasts is as accurate as possible and ideally with no false-positives.

Therefore, a set of conditions are required to start sharing a particular track with the teammates. As previously described, tracks with higher visibility rate have priority to be shared. Furthermore, it is required that the track age is higher than a certain threshold, it must have a minimum number of seen cycles and minimum visibility ratio, and must not exceed a maximum distance from the observing robot (to prevent false positive detections on higher distances).

Once a track is set to be shared, it will be shared until it is deleted or leaves the field.

3 Multi-Obstacle Tracking With Multiple Observations

In this Section, we present a methodology used to merge information from multiple agents to form a unified representation of the obstacles spread around the field.

In the context of the Middle-Size League, this is particularly important for the coach, which is a computer allowed to communicate with the robots, but not allowed to have any sensors attached. The objective of this coach is to give high-level coordination instructions for the team - strategy, formation, attitude, etc. Using the information shared by the robots, the coach is able to create a representation of the opponents position, which can be used to anticipate opponent gameplay, such as forward passes and set-plays.

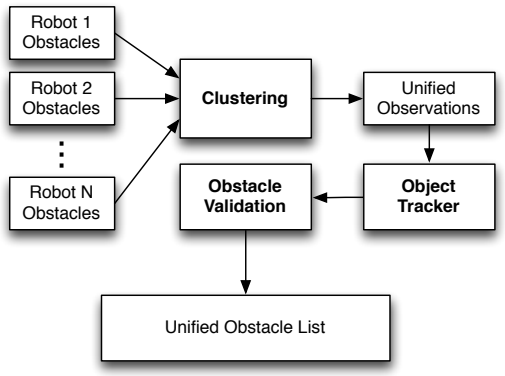


Figure 2: Summary of the tracking system using multiple agent shared observations

Figure 2 shows an overview of the global tracker. It considers the various agents shared obstacle tracks as observations. Although this shared information is not made of raw observations, but rather processed observations that have been associated together as a track and met the previously discussed criteria to be shared among the team, they can be considered observations for the purpose of creating this unified obstacle list.

3.1 Observation Clustering

The Hungarian Algorithm can be used to match observations with tracks. However, by solving a global minimization problem, it can not account for situations where there are multiple observations for the same object. Therefore, a clustering algorithm was implemented, based on the Constrained K-Means method [4], to take advantage of the background knowledge, that can be expressed as a set of instance-level constraints on the clustering process.

In the case of the MSL, the maximum size of the obstacles is limited by the rules (50×50cm), which implies that an obstacle that occupies the maximum allowed size can be perceived as a 70.7cm-wide obstacle (when seen from the diagonal). Despite this theoretical value, on this league, at the time of writing this paper, most teams opt for a triangular configuration on their platforms, meaning that size is not reached. Moreover, their sides do not measure less than 30cm.

Using this background knowledge, the *COP-Kmeans* method has been applied with the following constraints:

- if $width(centroid_i) > 0.7$m, split in two centroids
- if $distance(centroid_i, centroid_j) < 0.3$m, merge $centroid_i$ with $centr$

3.2 Applying the Object Tracker

The output of the previous clustering stage is a unified observation list, which is the input for the object tracker module. Its implementation was already described in Section 2.1.

3.3 Obstacle Validation

To avoid unreliable information, the unified tracks are subject to a validation. Once a track has been validated it will remain valid until it disappears.

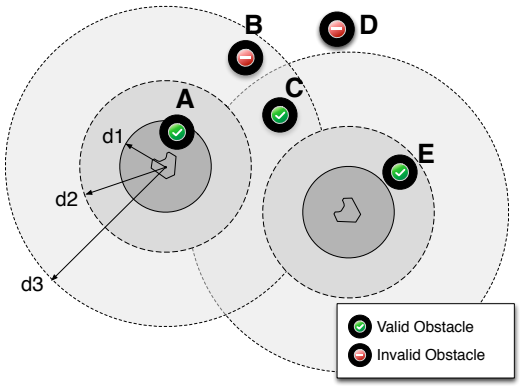


Figure 3: Example of the validation criteria using two robots. Here, three observations are validated and two do not pass the validation criteria. The image is not in scale.

Figure 3 shows an illustration of the implemented validation methodology. Essentially, the position of the team robots define three different zones:

- **Zone 1** - any position closer than d_1 from a team mate. In this zone, only tracks observed by the closest team mate robot can be validated. It is such a small distance that the closest robot must be able to see it directly. This prevents using detections of spurious obstacles close to team-mate positions.
- **Zone 2** - any position closer than d_2 from a team mate. Any track lying on one of these zones is validated.
- **Zone 3** - any position closer than d_3 from a team mate. A track lying on this zone requires at least two observing robots to be validated.
- Any tracked obstacle which distance from the closest team-mate is more than d_3 is not validated, since it is outside the maximum detection distance boundary.

4 Results and Discussion

In MSL, since 2016, teams are encouraged to supply their worldstate information during the matches (only for logging purposes), with the objective of providing the other team a reference to benchmark solutions after the game. In this case, knowing the exact location of the opponents after the match, allows us to use it as a groundtruth to match our obstacle detection algorithm.

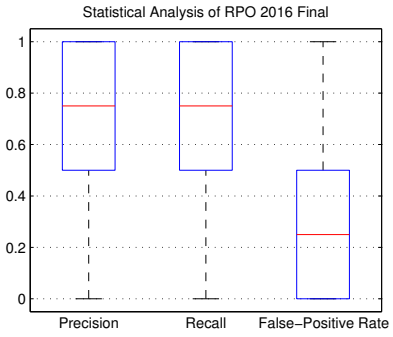


Figure 4: Detection Rate and False-Positive Rate results

Therefore, we implemented and tested this approach during the first half of the final of the RoboCup Portuguese Open 2016 and analysed our obstacle detection against the "groundtruth" provided by the other team. For this purpose, we only considered free-play situations, because in other situations, the referee may be inside the field repositioning the ball, and therefore inducing extra obstacles in the perception of the robots. Under these conditions, our dataset contains **1828 frames** sampled at **10 Hz**. Based on the false-positive rates and considering the predefined rules for validation on zones 1,2 and 3, the considered distance parameters were $d_1 = 1.0$m, $d_2 = 2.5$m and $d_3 = 5$m.

To benchmark the performance of this solution, we used three different metrics:

- **Precision** - percentage of correctly identified obstacles (over the detections) in each frame
- **Recall** - percentage of correctly identified obstacles (over the groundtruth) in each frame
- **False-Positive Rate** (FPR) - percentage of outliers in each frame

As Figure 4 shows, we obtained a median of:

- **Precision**: 75%
- **Recall**: 75%
- **False-Positive Rate**: 25%

Of course, the objective is always to maximise both precision and recall and to minimise the FPR. However, there is always a trade-off between between these two metrics, since it is always possible to increase distances d_1, d_2 and d_3, but not without sacrificing the FPR, because increasing those distances would mean to detect even more false-positives.

Experience and visual analysis tell us that these spurious detections occur mostly when our robots are moving fast (they can reach velocities of 4 meters per second) and the obstacles are far away from the detecting robots. As an improvement, the distances could vary with the robot velocity modulus.

It is also important to note that we are including false-positives that are created at the beginning of all this process - object detection - of which scope is outside this work. Moreover, by visual inspection we also noticed that sometimes the false-positive detections occur near our robots. Given our premises, it means that the robots sometimes wrongly identify obstacles in its vicinity. This can occur, for example, in cluttering situations, where strong shadows appear on the field. Since our obstacle detection relies mostly on colour (and the robots have to be mostly black), this sometimes creates false detections.

However, we further investigated the false-positives and the histogram in Figure 5 shows that we detected at most 8, but that most times the number stays between 0 and 2, which is acceptable, given the environment

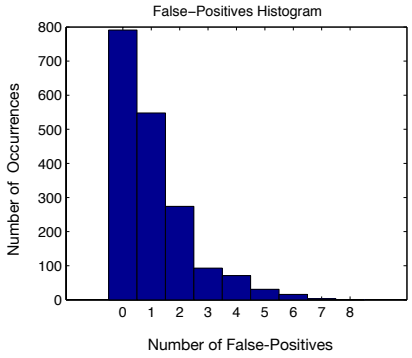


Figure 5: Histogram of the number of False-Positives

conditions of the test. In fact, in our previous approach (a naive merging algorithm, without validation), the robots could detect up to 20 obstacles in some situations, which would give us, at least, 15 false-positives. Unfortunately, there is no work from other teams which we can fairly compare these results with.

The evaluated metrics will allow us to further improve the algorithm, by optimizing the avaliable parameters, always with the objective of maximising the detection rate and minimising the False-Positive Rate.

5 Conclusion

In this paper we presented a solution to integrate information from several agents to formulate a unified world state representation. We started by discussing the implementation of an object tracking module and its application in tracking two different types of objects in a robotic soccer environment: obstacles and the ball.

We then moved to the presentation of a methodology to merge this information that is generated in (and shared by) different agents into an unified representation of the obstacles spread around the field.

This solution was implemented on a Middle-Size League agent integrator and tested during the RoboCup Portuguese Open 2016 competition.

The results show that a relatively high detection rate can be achieved, with some room for improvement concerning the false-positive rate. Nonetheless, this strategy played a major role in a number of situations, by allowing the team to act quicker, anticipate the opponent actions and even prevent dangerous situations like forward passes by covering an opponent from the ball.

References

[1] Wilfried Elmenreich. *Sensor Fusion in Time-Triggered Systems*. PhD thesis, Institut fur Technische Informatik, Vienna, Austria, 2002.

[2] L. Merino, F. Caballero, J.R.M.-d. Dios, and A. Ollero. Cooperative Fire Detection using Unmanned Aerial Vehicles. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 1884–1889, April 2005. doi: 10.1109/ROBOT.2005.1570388.

[3] A. Neves, J. Azevedo, N. Lau B. Cunha, J. Silva, F. Santos, G. Corrente, D. A. Martins, N. Figueiredo, A. Pereira, L. Almeida, L. S. Lopes, and P. Pedreiras. *CAMBADA soccer team: from robot architecture to multiagent coordination*, chapter 2, pages 19–45. I-Tech Education and Publishing, Vienna, Austria, January 2010.

[4] Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schrödl. Constrained k-means clustering with background knowledge. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 577–584, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1-55860-778-1.

[5] Michael Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, 2009.

Facial recognition based on image compression

Marco Henriques
marco.henriques@ua.pt
António J. R. Neves
http://sweet.ua.pt/an/
Armando J. Pinho
http://sweet.ua.pt/ap/

Departamento de Eletrónica e Telecomunicações
Universidade de Aveiro
Aveiro, Portugal

Abstract

Facial recognition has received an important attention in terms of research, especially in recent years, and can be considered as one of the best succeeded applications on image analysis and understanding. Proof of this are the several conferences and new articles that are published about the subject. The focus on this research is due to the large amount of applications that facial recognition can be related to.

Although there are many algorithms to perform facial recognition, many of them very precise, this problem is not completely solved mainly because of environment changes on the image's acquisition process.

The method proposed in this paper tries an innovative approach using similarity metrics obtained based on data compression, namely by the use of Finite Context Models to estimate the number of bits needed to encode an image of a subject, using a trained model from a database.

1 Introduction

Image compression is a way to reduce what is called the redundancy of image data, i.e., information that an image may have in excess, or it is not very relevant in terms of the image itself. On many cases, this relevance is related to the human's visual system: sometimes when compression is performed on an image, although some information was "thrown" away, it looks exactly the same for the human eye.

On other hand, there is a problem which has been studied since the beginning of computer vision. It is the facial recognition problem. It is quite easy for a human to recognize another human being, independently on weather conditions, distance (limited of course)...but is it that easy for a computer? Well, the task to confirm that a person is who claims to be, or to recognize (identify with a label/name) a certain person, can be a really hard task for a computer, because there are several obstacles which can be related to it (lighting, rotation or scale changes, for example).

Can these two subjects described above be related? In other words, can compression based algorithms be used to perform facial recognition?

This paper presents a study on the use of compression to verify similarity between images; in general, when a compression method is performed, one gets a number or metric at the end regarding some compression aspect. In this case, the output of the used compression algorithm is the number of bits to encode a certain image, therefore it will be used as a metric of similarity between images. In particular, this work focuses on the use of this similarity metric for the problem of face recognition.

2 Face recognition using image compression

2.1 A Similarity metric

There is work of some researchers, such as Kolmogorov, regarding the problem of defining a complexity measure of a string. The Kolmogorov complexity of a string x, K(x) is defined as the length of the shortest effective binary description of x. Broadly speaking, given a string of bits x, its Kolmogorov complexity is the minimum size of a program that produces x and stops. However, this measure of Kolmogorov complexity is not computable, so it needs to be approximated by other computable measures.

One of the choices to approximate the Kolmogorov complexity is through the usage of lossless compression, because such algorithms can be used (together with a decoder) to reconstruct the original data, and the amount of bits required to represent both the decoder and the bit-stream can be considered an approximation of the Kolmogorov complexity [2].

According to [1], a compression algorithm needs to be normal in order to be used to compute this metric, in other words, the algorithm must create a model while the compression is performed, accumulating knowledge of the data (finding dependencies, collect statistics).

A Finite Context Model provides, on an "on-line" symbol by symbol basis (the model is updated while the compression is performed, i.e., as the data is processed), an information measure that corresponds, in essence, to the number of bits required to represent a symbol, conditioned by the accumulated knowledge of past symbols.

2.2 Finite Context Models (FCM)

A finite-context model (Figure 1) collects statistical information of a data source, retrieving probabilities of occurrence of certain events. For every outcome, the Finite Context Model assigns probability estimates to the symbols from a certain alphabet $\mathcal{A}$. These estimates are calculated taking into account a conditioning context computed over a finite and fixed number k (the context size, which is a positive integer value) of the past k outcome symbols $c^t = x_{t-k+1},...,x_{t-1},x_t$.

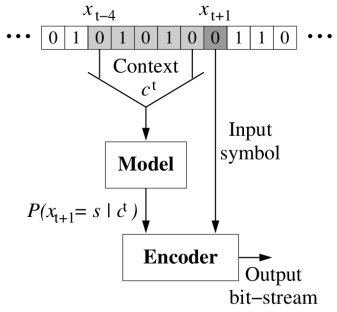


Figure 1: Finite Context Model.

The probability that the next outcome, X_{t+1} is $s \in \mathcal{A}$, followed by context c^t, is obtained by Equation 1

$$P(X_{t+1} = s|c^t) = \frac{N_s^t + \alpha}{\sum_{a\in\mathcal{A}} N_a^t + |\mathcal{A}|\alpha} \quad , \tag{1}$$

where $|\mathcal{A}|$ is the size of the alphabet, N_s^t represents the number of times that the source generated the symbol s having context c^t in the past and $\sum_{a\in\mathcal{A}} N_a^t$ is the total number of times the same context has appeared in the past. The α parameter is used to solve probability zero problems, corresponding to the first time a certain symbol appeared after a certain context; although this event remains unseen until a certain moment in time, there's still a certain probability of occurrence of that same event. On the subject of this work, its value is set to one, which is also known as the Laplace's Estimator.

$$B = -\sum_{t=0}^{N-1} \log_2 P(X_{t+1} = x_{(t+1)}|c^t) \ \ bit \ \ . \tag{2}$$

The data on this work are two dimensional arrays (images) and the pixels chosen as context are the spatially closest, because they are believed to correspond to the most recent past (in terms of statistics, they are more related with the central pixel than those further away on the image).

When a Finite Context Model is trained with a certain image, if one tries to encode the same image based on the trained model, the total number of bits will be smaller than the one to encode the image with no model. Plus, if the same model is used to process two different images, the number of bits will be smaller in the case of the most similar image (compared to the model). This information can be seen on Figure 2.

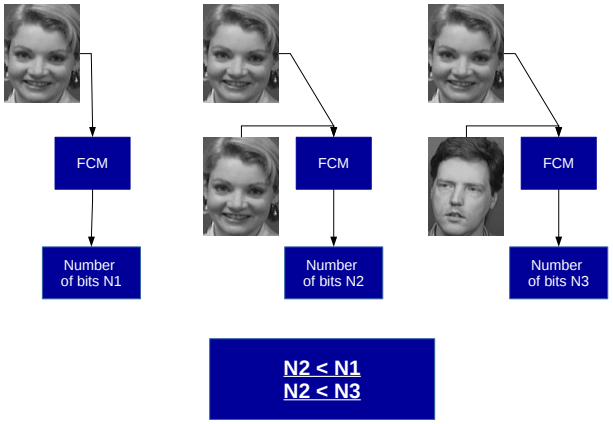


Figure 2: Principles of FCM on facial recognition.

3 Results

Two face databases were used as test subjects: "The Database of faces" (formerly the ORL Database of faces), and the Face Recognition Technology (FERET) database. Several tests were performed, based on the face recognition problem (number of subjects successfully/unsuccessfully recognized), for different context configurations and quantization levels, with and without changes on the images conditions. Some real-life tests were still performed.

Each database's subject has its own training images (9 on the ORL case, 3 on the FERET) and a test image. All models were created for each subject and the numbers of bits required to encode each of the test images with all trained models were calculated.

Therefore, and assuming that the minimum number of bits to encode an image of a subject is obtained with the trained model of the same subject, one can draw conclusions about the success of the face recognition. Thus, three different situations may occur:

- The minimum number of bits to encode the test image is obtained with the trained model of the same subject, and this number is within the limits (in terms of numbers of bits) of this model - subject successfully recognized;
- The minimum number of bits to encode the test image is not obtained with the trained model of the same subject, but this number is within the limits - subject recognized incorrectly (false positive);
- The minimum number of bits to encode the test image is outside the limits of the model - subject not recognized.

An example of the results achieved can be seen on Figure 3, where one can observe the three situations above mentioned (the highlighted points represent the minimum number of bits for each test image):

- Subject successfully recognized (green);
- Subject recognized incorrectly, false positive (red);
- Subject not recognized (yellow).

The information on the number of subjects successfully/unsuccessfully recognized can also be observed on Tables 1 and 2. As an example, Table 1 presents the recognition results using the ORL database.

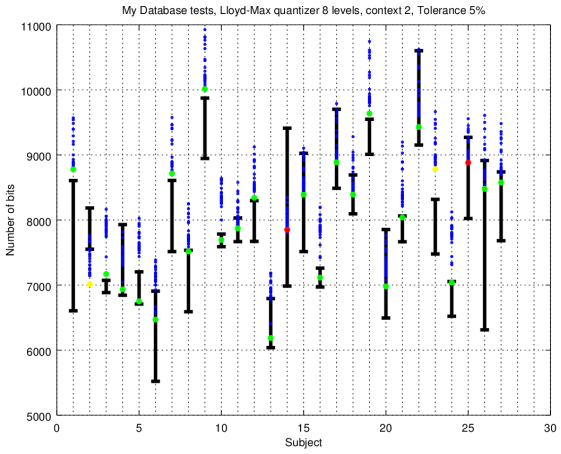


Figure 3: Recognition results: example.

Quantization Levels	Context	Positives	False Positives	Unrecognized
4	2	29	5	6
	4	29	4	7
	6	31	0	9
8	2	31	0	9
	4	23	0	17
16	2	25	3	12
	3	24	1	15

Table 1: Recognition results using several quantizations and contexts, ORL Database.

Several tests were also conducted in which there were changes in the image conditions, in order to study the versatility of the compression algorithm in those cases. These tests were achieved through light, rotation and scale changes on the test image. These image's alterations lead to changes to the number of bits needed when encoding the altered image with the same model. Table 2 presents the results of the lighting changes on the recognition process.

Lighting Factor	Positives	False Positives	Unrecognized
1	23	0	17
0.5	22	0	18
0.8	25	0	15
1.2	27	0	13
1.5	14	2	24

Table 2: Recognition results applying lighting changes on the test images.

4 Conclusions

The research presented in this paper had the main goal to develop and study a new possible and viable solution regarding the facial recognition problem, adding its own contribution to some research already performed on the subject.

This algorithm proved to work under constrained environments. However its performance was not so good on a non-constrained one. Some future research about this issue can improve the algorithm to be more flexible for other situations.

References

[1] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul Vitányi. The similarity metric. *Information Theory, IEEE Transactions on*, 50(12):3250–3264, 2004.

[2] Armando J Pinho and Paulo JSG 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.

Using Deep Machine Learning for Medical Image De-identification

Eriksson Monteiro, Carlos Costa, José Luís Oliveria
<eriksson.monteiro@ua.pt, carlos.cost@ua.pt, jlo@ua.pt>

Institute of Electronics and Informatics Engineering of Aveiro
Department of Electronics, Telecommunications and Informatics
University of Aveiro

Abstract

Clinical data sharing between healthcare institutions, and between practitioners is often hindered by privacy protection requirements. This problem is critical in collaborative scenarios where data sharing is fundamental for establishing a workflow among parties. The anonymization of patient information burn in medical images requires elaborate processes somewhat more complex than simple de-identification of textual information. In this paper, we propose an approach which applies image processing functions and deep machine-learning models to bring about an automatic system to anonymize medical images. For accessing the overall system quality, 500 processed images were manually inspected showing an anonymization rate of 89.2%. The tool can be accessed at https://bioinformatics.ua.pt/dicom/anonymizer.

1 Introduction

The protection of patient privacy is extremely important for any Electronic Healthcare Record (EHR) system. It becomes even more imperative in collaborative environments or clinical research trials [1]. There are textual data anonymizers but, these tools are inefficient in cases where patients' demographic information is present in the image pixel data such (e.g. ultrasounds with burned annotations), which may lead to the disclosure of patients' information. Therefore, medical image anonymizers should remove any sensitive information from both metadata and pixel data to be compliant with the HIPAA (Health Insurance Portability and Accountability Act) privacy rule which establishes standards for full de-identification of the patient [2].

This article proposes a methodology for performing medical image anonymization regarding pixel data. The methodology uses deep machine-learning techniques for the identification and removal of patients' demographic information in images. It proposes an end-to-end text recognition based on machine learning algorithms and it was developed using real data collected from a clinic of diagnostic imaging, namely ultrasounds produced by modalities of distinct manufacturers.

To promote use of the proposed methodology, we developed a public Web service to serve the clinical community, researchers and developers. In addition, we developed a Web application for end-users. The web application offers a user interface to access the available functionalities, an individual storage area with search functionality and a Web viewer for medical imaging.

A. *Medical Imaging*

Healthcare centres have quickly recognized the benefits of information technology (IT) in the management of medical information. Picture Archiving and Communication System (PACS) revolutionized radiology and, to some extent, medical practice [3]. It was necessary to create major infrastructures and workflows clearly defining how medical images are stored and accessed in a hospital network. PACS systems use the DICOM standard [3], which defines data formats, storage organization and communication protocols of digital medical imaging. This standard has emerged as a result of the appearance of equipment with the capacity to acquire, transfer and store imaging data, and the consequent need to standardize the communication processes of these devices on the network. DICOM files can support multiple types of medical information, such as images, waveforms, structured reports, etc.

B. *Medical Imaging Anonymization*

The use of medical images beyond the institutional border, for instance, to build phenotype-specific databases, for teaching and even for research purposes has been increasing. Thus, efficient tools to perform patients' de-identification are required. The anonymization process should remove or replace any information elements which may lead to patient identification. There is well-known protected health information (PHI) in the standard DICOM attributes [4].

Although earlier research outlined the importance of creating mechanisms to verify and remove PHI from DICOM metadata [5], there are still gaps regarding the removal of PHI annotations burned into the image pixels. Tools for this task already exist, but they are based on manual processes for identification of image areas containing PHI annotations [6]. Newhauser et al. [4] automated this process using optical character recognition (OCR) to detect text burned into the images. Nevertheless, the proposed algorithm removes all alphanumerical annotations. This is not desirable in some situations where some important annotations must be preserved, for instance, examination parameters and measurements in echocardiography.

The sharing of de-identified medical images retaining useful values for research is a challenge. Huang et. al [7] have proposed a method for preserving the privacy and security of patient medical records in this scenario. Another challenge is patient privacy protection in medical images containing facial features. Li et. al [8] pointed out the need of DICOM brain images defacing and proposed method to remove facial features without remove brain tissue to ensure correct brain images de-identification.

2 Methods

The proposed methodology for de-identification of pixel data in DICOM medical images follows a pipeline composed of 6 steps presented in Fig. 1. The first step consists of extracting PHI elements from the DICOM metadata that may be present in the pixel data. We extract the following sensitive DICOM attributes: patient name, ID, gender and accession number. This information is used to fill the set of words that need to be removed from image pixel data. In the next step of the pipeline, image preprocessing algorithms, such as adaptive bilateral filter and total-variation de-noising, are used to prepare the image for the following tasks – character recognition and classification. Once we obtain the preprocessed image, the object detection step is performed using functions to identify contours and detect objects through bounding rectangles that are subsequently normalized. Next, each object is classified using a machine learning-based OCR system that returns the identified character. After classifying the objects, the words present in the image are reconstructed, considering the objects' position and their classification. The next step detects which of the reconstructed words are present in the set of sensitive information.

Finally, sensitive words found are removed from the image pixel data by drawing a white rectangle over them. The next subsections will describe in more detail the most important steps of the proposed pipeline.

C. *Dataset and Model*

Our deep machine learning algorithm was develop using 62992 character samples extracted from the *73K Character* dataset [9]. Each image in the dataset is labeled with the respective alphanumerical present in it (i.e. 0-9, a-z and A-Z).

Extra noisy data were artificially generated by performing morphological transformations of the images (i.e. erosion, dilation) and also applying a convolution (linear shifts of 1 pixel in each direction) to the samples in the training data. As a result, the final dataset comprised 944880 labeled samples. We split the dataset between a train set and a test set through stratified sampling where we took 90% of the samples to build the train set and the remaining 10% of the samples was left for the test set.

A deep convolutional neural network (CNN) [10] was trained to recognize characters. CNN is a variant of Multilayer Perceptron (MLP) network and is one of many deep learning algorithms [11]. This algorithm exploits spatially-local correlation and it has been found effective in general image classification tasks [10].

The deep convolution neural network created was composed of several layers, combining convolution layers and subsampling layers (max-pooling) and in the end we attached two dense layers (Fig. 2).

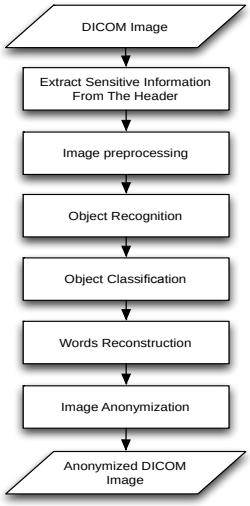


Fig. 1 - Anonymization pipeline for pixel data in DICOM images

The input image is reshaped to an image with a pixel area of 32x32, then the image is processed in a convolution layer generating 64 images of 20x20. The network also comprises subsampling layers, which reduce the complexity of the model, by applying max-pooling filters on input data. For example, we have in the second layer a max-pool layer to reduce the images size 20x20 to 10x10. Subsequently, we added one more convolution and max-pooling layers, then, there is a dense layer to convert the data to a feature vector with 256 features. Finally, the dense layer is connected to the output layer, which has 36 neurons representing each class (alphanumerical character).

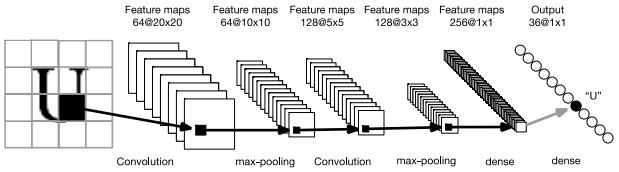


Fig. 2 - Deep convolution neural network layers and feature maps

3 Results

D. *Deep machine-learning OCR*

We used a test set composed of 94488 samples to test the model performance in terms of precision, recall and F1-score. In the end, we obtained the results for the measures, where we obtained 96.66% precision, 96.30% recall and 96.47% F1-score.

E. *Anonymization Pipeline Evaluation*

This section, moves on to evaluate the whole anonymization pipeline where we used the deep CNN model for object classification. An automated process was implemented to anonymize a set of studies extracted from a real-world PACS archive. In result assessment process, we had to verify whether all sensitive information was removed from the resulting images, marking them as correctly anonymized, or as anonymization errors if we left one or more pieces of sensitive information visible in the image. We processed 500 ultrasound studies with different patient names and, then, we visually inspected the resulting images to measure performance Fig. 3.

Thus, for each study we observed which one had all sensitive words removed (anonymized), which one failed to remove (not anonymized) and which one had some words / image region mistakenly removed (mistake). Accordingly, we obtained a 89.2% success rate (anonymized), 10.8% of not anonymized images and, 1.0 % of images with some regions mistakenly removed as can be observed in Fig. 3.

4 Conclusions

As a result, we developed a machine learning-based system with a reliable anonymization success rate, and which consequently, may have several applications, such as in improving current methods that only consider DICOM metadata.

Summarily, our methodology presented promising results and it may represent added value in automatic patient de-identification in medical imaging (Fig. 4), offering an agile solution for use cases of medical image-sharing among institutions and users. Moreover, the system may be used in real world environments for supporting automated de-identification processes, since it identifies the cases where the anonymization is not fully achieved, through the inspection of sensitive tokens extracted from DICOM metadata.

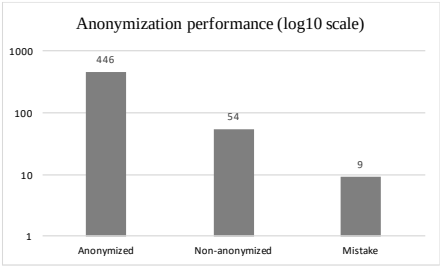


Fig. 3 - Overall system performance

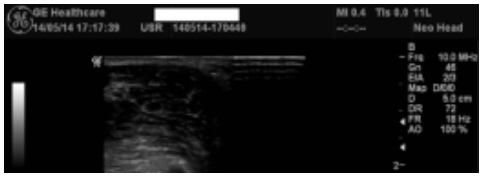


Fig. 4 - Anonymized image

References

[1] A. Suinesiaputra, P. Medrano-Gracia, B. R. Cowan, and A. A. Young, "Big Heart Data: Advancing Health Informatics Through Data Sharing in Cardiovascular Imaging," *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 4, pp. 1283–1290, Jul. 2015.

[2] J. B. Freymann, J. S. Kirby, J. H. Perry, D. A. Clunie, and C. C. Jaffe, "Image data sharing for biomedical research--meeting HIPAA requirements for De-identification.," *J. Digit. Imaging*, vol. 25, no. 1, pp. 14–24, Feb. 2012.

[3] H. K. Huang, *PACS and Imaging informations: basic principles and applications*. Wiley-Blackwell, 2004.

[4] W. Newhauser, T. Jones, S. Swerdloff, W. Newhauser, M. Cilia, R. Carver, A. Halloran, and R. Zhang, "Anonymization of DICOM electronic medical records for radiation therapy," *Comput. Biol. Med.*, vol. 53, pp. 134–140, 2014.

[5] D. Rodríguez González, T. Carpenter, J. I. van Hemert, and J. Wardlaw, "An open source toolkit for medical imaging de-identification," *Eur. Radiol.*, vol. 20, no. 8, pp. 1896–1904, Aug. 2010.

[6] D. Clunie, "How to use DoseUtility TM," *PixelMed Publishing*. [Online]. Available: http://www.dclunie.com/pixelmed/software/webstart/DoseUtilityUsage.html. [Accessed: 26-Jul-2016].

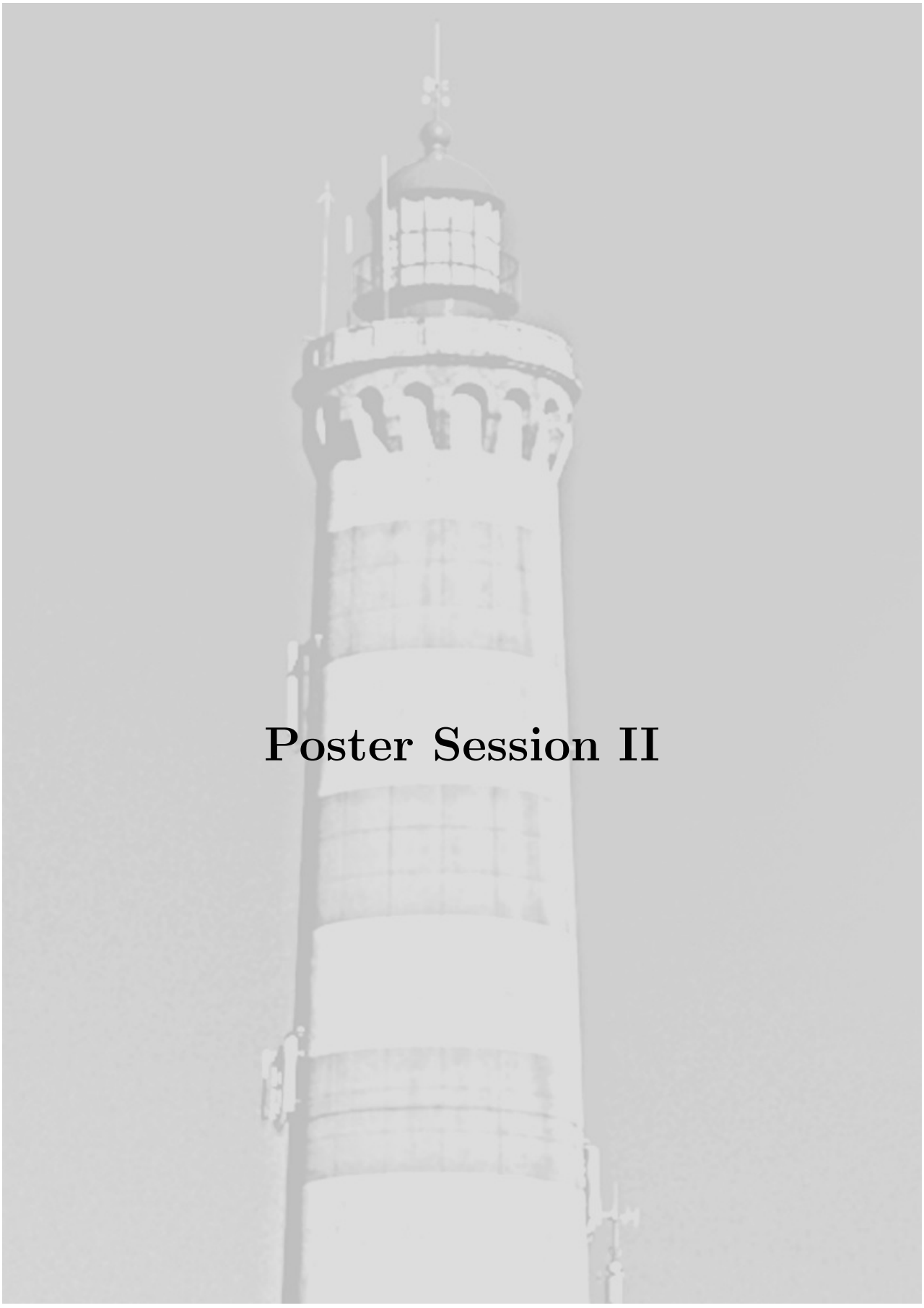
[7] L.-C. Huang, H.-C. Chu, C.-Y. Lien, C.-H. Hsiao, and T. Kao, "Privacy preservation and information security protection for patients' portable electronic health records.," *Comput. Biol. Med.*, vol. 39, no. 9, pp. 743–50, Sep. 2009.

[8] L. Li and J. Z. Wang, "DDIT - A Tool for DICOM Brain Images De-Identification," in *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, 2011, pp. 1–4.

[9] T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images," *Proc. Int. Conf. Comput. Vis. Theory Appl.*, 2009.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[11] Y. Bengio, "Learning Deep Architectures for AI," *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.



Poster Session II

Parametric Modeling of Breast Data Using Free Form Deformation

Hooshiar Zolfagharnasab
hooshiar.h.z@ieee.org
Jaime S. Cardoso
jaime.cardoso@inesctec.pt
Hélder P. Oliveira
helder.f.oliveira@inesctec.pt

Faculdade de Engenharia, Universidade do Porto
Campus da FEUP, Rua Doutor Roberto Frias, 4200-465
Porto, Portugal
INESC TEC
Campus da FEUP, Rua Doutor Roberto Frias, 4200-464
Porto, Portugal

Abstract

Nowadays, breast cancer has become the second most common cancer amongst females. Since breast is a symbol of feminine, any imposed deformations during surgical procedures affect the patients' quality of life. Therefore, alongside the successful removal of tumor, aesthetic shape of breast should be also considered. These objectives can be achieved by using a planning tool based on parametric model of patient's own breast. The application of a planning tool not only improves surgeons' skills to perform surgeries with better cosmetic outcomes, but also increases the interaction between surgeons and patients during necessary decision.

Studying a methodology of parametric modeling called Free-Form Deformation (FFD), two simplified versions of FFD are proposed to increase model similarity to input data and decrease required fitting time. Quantitative analysis indicate that the proposed modifications fulfilled both mentioned objectives to generate parametric models.

1 Introduction

Accounting for almost 29% of all newly diagnosed cancers in 2015, breast cancer is one of the two most frequent cancers among women [6]. As long as surgery is assumed as the most common treatment, breast deformation is taken place consequently after both surgical types; mastectomy in which the whole breast is removed, and lumpectomy, also called as Breast Cancer Conservative Treatment (BCCT), in which the tumor together with a thin layer of healthy surrounding tissue are removed [4].

Despite differences between the amount of removed tissue, survival rate of both surgeries are almost the same; therefore, since BCCT requires less tissue to be removed, final aesthetic outcome is expected to be more satisfactory to the patients [2].

Therefore, providing a tool to increase the interaction between patients and surgeons can be an interesting framework to assist surgeons in planning surgeries with better cosmetic outcomes. Besides, equipping aforementioned tool with mathematical formulated models, can provide a framework to generate different post-surgical breast shapes.

2 Background

Parametric modeling is the process of transforming specific data to mathematical models. Highlighting its application in human body, Weiss *et al*. [7] fitted the parameters of SPACE (Shape Completion and Animation for PEople) predefined models to depth data and image silhouettes of human body. In their work, they fitted each scanned body data to the closest predefined model.

Inspiring from the work done by Oliveira *et al*. [4], the aim is defined by generating 3D parametric models of breasts of the patients who undergone BCCT. Note that the views used for 3D reconstruction are obtained by Microsoft Kinect as a low-cost 3D scanning device.

2.1 Free-Form Deformation

The methodology proposed in [1] uses a two-step algorithm to perform parametric fitting. Beginning with the first step, a superellipsoid (as a superquadric model) is fitted to input data through changing its parameters in order to minimize the Euclidean distance between the input data and the superellipsoid. The second step, where the fitted superellipsoid is deformed, is initialed by defining a set of 3D parallelepipedic grid of control points around the fitted superellipsoid [8].

Containing $(l+1)(m+1)(n+1)$ points, the grid is linked to the fitted superellipsoid using a tensor product of tri-variants Bernstein's polynomials [8]:

$$X = \sum_{i=0}^{l}\sum_{j=0}^{m}\sum_{k=0}^{n} C_l^i C_m^j C_n^k (1-s)^{l-i} s^i (1-t)^{m-j} t^j . (1-u)^{n-k} u^k P_{ijk} \quad (1)$$

The parametric model is defined as X, and P is a matrix containing the points of the control grid. Besides, s, t, and u are used to denote local coordinate of each point of parametric model regarding to its corresponding control point. The Equation 1 can be redefined in two parts the summations:

$$X = BP \quad (2)$$

where B and P are called the deformation matrix and control points respectively. Considering δX as the displacement field between input data and superellipsoid, Equation 2 can be rewritten regarding to linear equation system:

$$\delta X = B\delta P \quad (3)$$

The superellipsoid is then deformed through relocating the control points of the grid iteratively to be approached to the input data.

3 Moving Towards Better and Faster Breast Fitting using FFD

Analyzing aforementioned FFD approach with 3D data which contain an open side [8], demonstrated an issue called Redundant Bent Layer (RBL) where the performance of modeling is affected negatively by increasing distance amongst the model and input data.

Geometrical analysis of the parametric modeling reveals that RBL occurs when both closed sides of the initial model (both hemispheres of the fitted superellipsoid) are pulled to the closed side of input data. Since the control points are arranged in a 3D formation, the two pulled sides of the initial model cannot be overlaid completely. Therefore, a gap is imposed between two sides of deformed model which produces wrong fitted surface that increases the distance between parametric model and input data. In order to eliminate the RBL on the parametric model, two solutions can be taken into consideration by modifying either the initial model, or the arrangement of the control points.

3.1 Modifying the Initial Model

As discussed before, the geometrical properties of the initial model reinforce the RBL issue. Therefore, replacing it with an open-side object can be a possible solution. In this paper, it is proposed to replace current initial model (superellipsoid) with two new ones: a finite boundary plane, and a superquadric model based on a superparaboloid similar to the one proposed in [5].

The first solution suggests to use a finite boundary plane to fulfill the requirements of using an open-sidedness. On the other hand, the second solution proposes to use a superparaboloid (introduced in [5]) in order to decrease the number of required iterations to have faster approach.

In both solutions, it is recommended to set them up orthogonal to the largest principle direction of data. This assures that the model is initiated in a correct place where can be overlaid to the input data completely. For this purpose we used a Principle Component Analysis (PCA) approach.

3.2 Modifying the Arrangement of Control Points

Originally, in [1], it was suggested to use control points arranged in a 3D grid. However, the input data (breast) are presented as an open-side object. Therefore, whilst modeling, the control points located in the open-side of the initial model not only provide a circumstance in which RBL can be emerged, but also impose higher computation (and time) cost due to their participation in the calculation of point's new location. Therefore, the second proposal suggests to remove ineffective control points by reducing the dimensions of the control grid from 3D to 2D.

Dimension reduction should be carried out with regard to the presentation of input data. Common methodologies to perform PCA can lead to obtain the best removal candidate. Assuming the third dimension as the candidate of removal, the proposed dimension reduction of control points simplifies the nested summations used to relocate control points in Equation 1:

$$X_{new} = \sum_{i=0}^{l} \sum_{j=0}^{m} C_l^i C_m^j (1-s)^{l-i} s^i (1-t)^{m-j} t^j . P_{ij} \quad (4)$$

X_{new} denotes the points of the parametric model. The less control points are considered in the computation, the less time is required for fitting.

4 Implementation and Results

The original methodology discussed in [1] together with proposed methodologies were implemented using C++, and evaluated on a 3.40 GHz machine equipped with 8 GB of memory.

Iterative modality of the studied algorithms requires to define a stop criterion. Such requirement is defined in relation to the Euclidean distance between the models being generated in each two consecutive steps.

Two metrics are considered for the evaluation of the explained methodologies; distance error and number of iterations required for fitting. Distance error stands for the average of Euclidean distance between the input data and the generated parametric model. Not only Euclidean, but also Hausdorff distance is calculated. The smaller the distance error is, the more similar the two sets will be. Besides, number of iterations required to reach to final deformation is also a key advantage in comparisons. Expectedly, approaching with less iterations is more preferable.

The evaluated dataset have been obtained by scanning 36 patients using Microsoft Kinect as a low-cost scanner. Afterward, 3D model of patients are reconstructed via the algorithm proposed in [3].

Table 1 presents the results obtained by the different methodologies. Bi-directional distances are evaluated due to difference amongst the number of points of the two compared clouds. Also, the average number of required iterations to reach to stop criterion are reported. Beside the numerical analysis, visual comparisons are depicted in Figure 1.

A brief look to the Table 1 and the requirements of the fitting stage reveals that the methodologies based on FFD superellipsoid and plane generate more precise parametric models than superparaboloid. Reportedly, the least Euclidean error from parametric model to input data was 1.21mm which is obtained by the superellipsoid and a 2D FFD. The methodology of using a plane with a 2D FFD takes the second rank with error of 1.25mm.

Considering Model to Ground-truth Euclidean distance, the suggested improvement of 2D arrangement of control points surpasses other method since it eliminates the RBL phenomena. With a small gap, the methodology of using plane with 2D FFD stands in the second rank since using plane as the initial model cannot present the boundaries of breast better than superellipsoid.

Comparing number of iterations, both superparaboloid methodologies (with 2D and 3D FFD) are ranked in the first and second places, that is because of the similarity between initial model and breast data. In the next place, plane with 2D FFD stands with less than 4 iterations.

5 Conclusion

Mentioning the importance of a parametric modeling in a planning tool, methodologies of FFD have been studied in this paper and two improvements were proposed to enhance it; improvement of the initial model and modification of control points arrangements. Quantitative analysis indicated the proposed approaches improve the performance of FFD methodology by decreasing the average distance error. Possible future work will be concentrated to generate parametric models with less shrinkage which leads to less distance error.

Acknowledgment

This work was funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF), and also by Fundação para a Ciência e a Tecnologia (FCT) within PhD grant number SFRH/BD/97698/2013.

References

[1] 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.

[2] M. J. Cardoso, H. Oliveira, and J. Cardoso. Assessing cosmetic results after breast conserving surgery. *Journal of surgical oncology*, 110(1):37–44, 2014.

[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] 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.

[5] 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.

[6] American Cancer Society. American cancer society: Breast cancer facts and figures 2015-2016. *American Cancer Society (ACS).*, 2015.

[7] A. Weiss, D. Hirshberg, and M. J. Black. Home 3d body scans from noisy image and range data. In *in IEEE International Conference on Computer Vision*, pages 1951–1958. IEEE, 2011.

[8] H. Zolfagharnasab, J. S. Cardoso, and H. P. Oliveira. A 3d parametric model for breast data. In *Proc. the 21th edition of the Portuguese Conference on Pattern Recognition (RECPAD)*, pages 40–41, 2015.

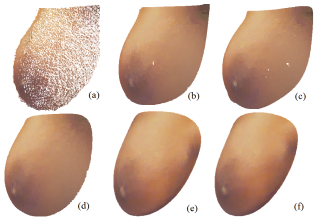


Figure 1: Visual comparison of performed experiments on same input data; (a) original patient breast; (b) Plane + FFD (2D); (c) Superellipsoid + FFD (2D); (d) Superellipsoid + FFD (3D); (e) Superparaboloid + FFD (2D); (f) Superparaboloid + FFD (3D)

Table 1: Reported results to compare proposed methodologies of FFD. μ and σ are average and stan deviation of distances from generated parametric model (M) to groundtruth (GT) breasts, respectiv

Methodology		Euclidean (M→GT) (mm)	Euclidean (GT→M) (mm)	Hausdorff (M→GT) (mm)	Hausdorff (GT→M) (mm)	Mean No. Itr
Superparaboloid + FFD (3D)	μ	1.57	2.28	11.02	26.28	3.60
	σ	0.25	0.48	3.93	7.84	
Superparaboloid + FFD (2D)	μ	1.60	3.03	7.14	29.62	3.2
	σ	0.25	0.81	1.76	7.62	
Superellipsoid + FFD (3D)	μ	1.31	1.70	7.63	13.43	10.42
	σ	0.08	0.25	1.52	4.83	
Superellipsoid + FFD (2D)	μ	1.21	2.71	5.62	24.27	5.33
	σ	0.11	0.52	1.33	7.49	
Plane (2D) + FFD (3D)	μ	1.25	2.78	6.27	23.67	3.83
	σ	0.11	0.52	1.71	8.23	

Psychophysiology assessment tool using Virtual Reality - Case Study

Bernardo Marques[1]
bernardo.marques@ua.pt
Susana Brás[1]
susana.bras@ua.pt
Sandra C. Soares[2]
sandra.soares@ua.pt
José M. Fernandes[1]
jfernan@ua.pt

[1] IEETA, DETI
Universidade de Aveiro
Aveiro, Portugal

[2] Department of Education and Psychology
Universidade de Aveiro
Aveiro, Portugal

Abstract

The treatment of specific phobias is gradually changing and there are currently works, which report the Virtual Reality exposure to be as effective, when compared to in vivo exposure. We propose Veracity, a low cost, portable and easy to deploy system for the monitoring of individuals. The selected case study was the phobias, more specifically, spider phobia. The system implementation allows the acquisition of multimodal information, when the individuals are confronted with a scenario that uses hand gesture combined with Virtual Reality, in order to force stimulus and capture the related data (ECG, HR, VIDEO, Screenshots, etc.) using external resources and a smartphone application.
In a one participant trial, with Veracity, it was possible to observe the alterations induced in the HR median and dispersion with and without the spiders - the phobic stimulus.

1 Introduction

Anxiety disorders affect many individuals, conditioning their daily life routines [5, 6]. Specific phobia is one example of an anxiety disorder, which is an irrational fear towards an object, or situation [1]. Phobics felt a distorted reality, and usually try to avoid the phobic element, which will only intensify the problem.
The evolution of technology and the miniaturization brought to the foreground not only allow the development of portable solutions for the assessment of psychological and behavior but also new possibility to mimic the real world likes Virtual Reality (VR) outside of the laboratory setting. Recent studies, describe Virtual Reality exposure as effective, when compared to in vivo exposure [2, 3], with the benefit of being less aggressive [7] to the phobic individual.
In this work, we propose Veracity, an affordable and portable system, which relies on VR to present more ecological and virtual stimuli to phobic individuals while monitoring their physiological and behaviour response. We used spider phobias as the main case study of our system, however design of our solution was though in order to allow its adaptation to other scenarios. While presenting the VR stimuli, using a game divided in a set of increasing difficulty levels, Veracity allows the interaction with the virtual environment through hand gesture. Simultaneously the individual's reaction is acquired (ECG, HR, RR, VIDEO, Screenshots, 3D objects tracking, etc.) using external resources and two smartphone applications. Veracity also supports data management for post processing integrated with the cloud. Veracity is quick to set up and allow outside of laboratory data collection and stimuli exposure, which allow more natural reactions of the individuals with specific phobias.

2 Methods

Veracity system combined with the "catch the spiders" game gives response to the need to force specific stimulus, in our case, regarding the spider phobia case study. Veracity can be quickly set up and presented in a daily life environment i.e. outside the laboratory, allowing more natural reactions from the phobic individuals. Its design allows adaptation to other scenarios as soon as clear scenario protocol is well defined i.e. stimulus, interaction paradigm and experimental protocol.
The current system (Figure 1) was developed and deployed in a MacBook Air, 13-inch, 1,4 GHz Intel Core i5, 8GB DDR3, Intel HD Graphics 5000, running OS X Yosemite 10.10.3 (1), a Samsung S3 GT-I9300, running Android 4.3 (3) and an action sport camera like Sony HDR AS30V (5), which besides providing a wide field of view, also allow to acquire the video from 25 Hz at 1080p to 120 Hz and 720p. The main components

Figure 1: Veracity current system deployment: (1) Laptop, (2) Leap motion, (3) Mobile device, (4) ECG monitoring solution, (5) Video camera.

of the system are the Virtual Reality Environment, User Interaction Device, Media Recorder, ECG/Heart Rate Measuring and Data Manager.
Having the biosignal data regarding the ECG and HR, combined with the information of the duration of each level in the game stored, it is possible to use Matlab software to build a quick preview of the collected data like it is illustrated in Figure 2.

2.1 Statistical Analysis

Our main objective was to assess if using Veracity, the physiological response (in this case heart rate) was modulated by the sequence of the adaptation and active levels - the later presenting a progressive exposure to spiders. Our hypothesis is based on the assumption that there will be differences induced by the spider (active levels) in relation to adaptation levels (no spiders).
For data analysis the Matlab software was used, for HR medians comparison between levels a Wilcoxon signed rank test was performed. The data dispersion was also important to infer the impact of the stimulus in the participant, in that case a Two-sample F-test was implemented, testing the alternative hypothesis that the population variance of x is lower than that of y. Both approaches evaluate multiple comparisons, in order to evaluate differences between game levels considering an alpha of 0.05. Therefore, and in order to accomplishes type I errors a post hoc correction should be performed, in that case we selected the Bonferroni correction.

3 Results

Aiming to test the proper functioning of the multiple modules of our system, a test was performed to the system with one participant (female, 35 years old). Besides targeting the initial hypothesis, the test allowed the evaluation of the overall data acquisition workflow (e.g. data collecting, online monitoring and event tagging).
The participant was asked to fill out the Fear of Spiders Questionnaire [4]

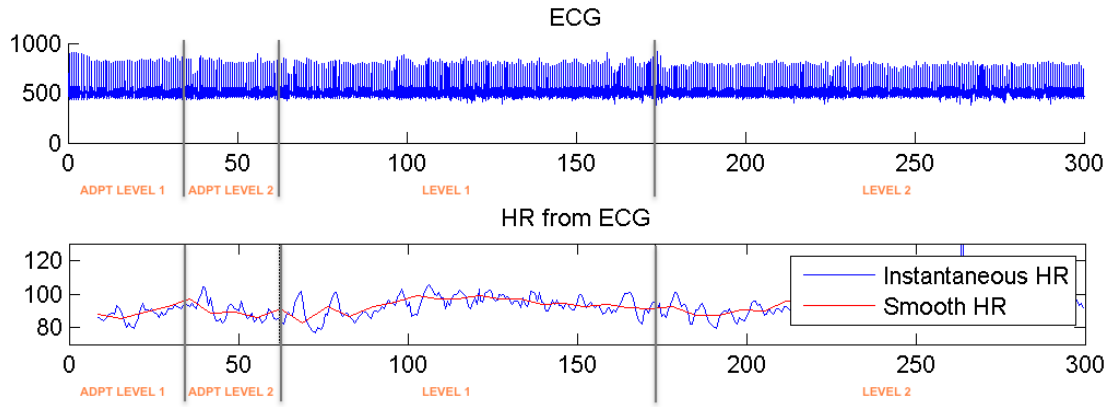


Figure 2: Sample collection: the synchronous acquisition of the several data modalities. The transitions between levels (vertical lines) are mapped into both ECG (in mV) and corresponding HR (in bpm), providing not only visual help separating the collected data of each level in the game, but also a way to quickly infer the time spent in each one. In both graphs, the vertical axis represent the biomedical data values and the horizontal axis correspond to seconds. The smooth HR was calculated based on moving avearge filter using a 10 beats window.

with the goal to identify possible spider phobics. Based on the questionnaire score, the participant was considered a probable phobic. However, no formal diagnosis of fear was performed. Figure 3, illustrates the obtained results, according to every level in the game. As such, it is possible to infer that the participant, can be considered the "perfect subject" in this study, taking into consideration that there is an increase on the median values, over the previous levels values, ($p<0.003$). In the Adaptation Level 1 and 2, the HR values start lower, but, when the spider stimulus is introduced in Level 1, and so on, there is a considerable increase in the HR values. Furthermore, it is also possible to verify that there is an increase in the HR dispersion values in the levels with the spider stimulus, when compared with the adaptation levels, which was again statistically verify ($p<0.003$).

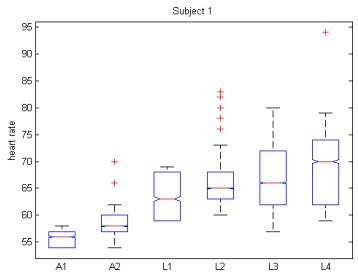


Figure 3: Boxplot representing the smooth-heart rate values (vertical axis) over the adaptation and active game levels (horizontal axis), regarding the participant overall study. A1/ A2 are the adaptation levels 1 and 2. L1/ L2/ L3/ L4 are the active levels, with phobic stimulus exposure.

Overall, the statistical analysis proved that there is an increase in the variance values of the participant HR values, regarding the levels with the spider stimulus, when compared with the adaptation levels.

4 Conclusion

Veracity is a generic system, based on low cost off the shelf components that provides an end-to-end solution, able to allow a Semi-immersive Virtual Reality scenario experience. The system allows phobic stimulus presentation and non-intrusive physiological monitoring. Veracity was developed under the goal of presenting a game divided in a set of increasing difficulty levels, which allow the phobic individual to interact with specific stimulus, in our case, spiders.

The system should be tested in a larger database, in order to verify the system efficiency in the stimulus reaction by the participant.

5 Acknowledgment

We would like to thank the volunteer participation on the study.
This work was supported by the European Regional Development Fund (FEDER) and FSE through the COMPETE programme and by the Portuguese Government through FCT - Foundation for Science and Technology, in the scope of the projects UID/CEC/00127/2013 (IEETA/UA), and CMUP-ERI/FIA/ 0031/2013 (VR2Market), PTDC/EEI-SII/6608/2014. S. Brás acknowledges the Postdoc Grant from FCT, ref. SFRH/BPD/92342/ 2013.

References

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, (DSM-5®)*. American Psychiatric Pub, 2013. ISBN 0890425574.

[2] Cristina Botella, M Ángeles Pérez-Ara, Juana Bretón-López, Soledad Quero, Azucena García-Palacios, and Rosa María Baños. In Vivo versus Augmented Reality Exposure in the Treatment of Small Animal Phobia: A Randomized Controlled Trial. *PloS one*, 11(2):e0148237, 2016. ISSN 1932-6203.

[3] Azucena Garcia-Palacios, Cristina Botella, H Hoffman, and Sonia Fabregat. Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & behavior*, 10(5):722–724, 2007. ISSN 1094-9313.

[4] Rafael Klorman, Theodore C Weerts, James E Hastings, Barbara G Melamed, and Peter J Lang. Psychometric description of some specific-fear questionnaires. *Behavior Therapy*, 5(3):401–409, 1974. ISSN 0005-7894.

[5] Richard T LeBeau, Daniel Glenn, Betty Liao, HansâĂŘUlrich Wittchen, Katja BeesdoâĂŘBaum, Thomas Ollendick, and Michelle G Craske. Specific phobia: a review of DSMâĂŘIV specific phobia and preliminary recommendations for DSMâĂŘV. *Depression and anxiety*, 27(2):148–167, 2010. ISSN 1520-6394.

[6] Arne Öhman and Susan Mineka. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review*, 108(3):483, 2001. ISSN 1939-1471.

[7] Bunmi O Olatunji, Brett J Deacon, and Jonathan S Abramowitz. The cruelest cure? Ethical issues in the implementation of exposure-based treatments. *Cognitive and Behavioral Practice*, 16(2):172–180, 2009. ISSN 1077-7229.

Facial Key-Points Detection using a Convolutional Encoder-decoder Model

Pedro M. Ferreira[1, 2]
pmmf@inesctec.pt
Jaime S. Cardoso[1, 2]
jaime.cardoso@inesctec.pt
Ana Rebelo[1]
arebelo@inesctec.pt

[1] INESC TEC,
Campus da FEUP, Rua Dr. Roberto Frias, 4200 - 465,
Porto, Portugal

[2] Faculdade de Engenharia da Universidade do Porto,
Rua Dr.Roberto Frias, s/n, 4200-465,
Porto, Portugal

Abstract

This paper addresses the problem of detecting key-points in face images. This is a very interesting topic, since it can be used as building block in several applications, especially in face recognition and biometrics. In this regard, several methods for facial key-point detection were implemented and compared, namely a mean patch searching algorithm, a convolutional neural network (CNN) model, and a convolutional encoder-decoder model. Experimental results show that, although the CNN model achieved the best results, the proposed convolutional encoder-decoder model has a great potential to be used for facial key-points detection.

1 Introduction

Facial key-points detection is an appealing topic with a great interest in the research community, since it can be used as building block in several computer vision/machine learning problems, such as biometrics, facial expressiveness analysis and face detection and recognition [4]. However, the task of identifying the facial key-points positions in an image is very challenging because facial features may vary greatly from one image to another, mainly due to the inter-individual's generic appearance variability. Moreover, facial features can also be affected by several physical and psychological factors such as position, viewing angle, occlusions, illumination conditions, contrast, and facial expressiveness (see Figure 1).

Previous methods for facial key-points detection can be roughly classified into two main categories: texture-based and shape-based methods [3]. Texture-based methods model the local texture by considering the grey-level values in a small neighbourhood around a given key-point. Typical texture-based methods include feature extraction methods, such as Gabor features. Shape-based methods consider all facial key-points as a shape instead of detecting each one individually. Typical shape-based methods include detectors based on active shape or active appearance models. Recently, with the trend of deep learning, several deep structures have been proposed and designed for this task [2].

The purpose of this paper, motivated by the online Kaggle[1] competition, is to predict a total of 15 facial key-points as illustrated in Figure 1. To accomplish this purpose, two main methodologies for facial key-point detection were explored:

1. Mean patch searching methods;
2. Deep leaning methods.

The first methodology is a simple mean patch searching algorithm with correlation scoring to predict the key-points positions. Regarding the deep learning strategies, a single CNN, to regress the spatial key-points positions across the image, was first considered. Moreover, we propose the usage of a convolutional encoder-decoder model to this problem. It consists in a deep net with a downsampling step followed by an upsampling step. The convolutional encoder-decoder model is used for regression of a density map, which is a probability density function (*pdf*) of the location of the key-points in the image. Once the *pdf* is estimated, the facial key-points positions are given by the local maxima of the density map. The big advantage of the convolutional encoder-decoder model is that we do not need to specify *a priori* the number of key-points to be predicted, as it happens with the traditional CNN's models in which the number of units in the output layer have to match the number of key-points. This could be very interesting specially if we have to deal with partially occluded facial images, in which some key-points are not visible and hence should not be detected.

[1] https://www.kaggle.com/c/facial-keypoints-detection/

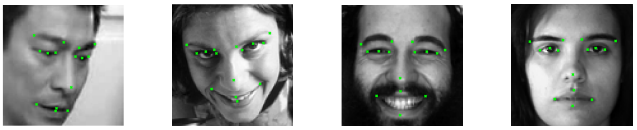


Figure 1: Samples of face images from Kaggle competition with 15 key-points superimposed.

2 Methodology

2.1 Mean patch searching

The mean patch searching approach is a relatively simple supervised learning algorithm, in which two main variables have to be learned from the training data for each key-point, namely their corresponding mean patch and average position.

Given n training images along with the corresponding true coordinates $(x,y)^j, j = 1,...,k$, of all k key-points. The underlying idea is to extract a square patch (with $patch\_size \times patch\_size$ pixels) around each key-point in each image, and then average the result. The mean patch of each key-point is computed as an entry-wise mean of the n patches obtained for each key-point. Formally, let $P_i^j \in \mathbf{R}^{(patch\_size \times patch\_size)}$ be the matrix representing the i^{th} patch $(i = 1,...,n)$ of the j^{th} key point $(j = 1,...,k)$, then the mean patch $\overline{P^j}$ of each key point is given by:

$$\overline{P^j} = \frac{1}{n}\sum_{i=1}^{n} P_i^j, \forall j \in 1,...,k \quad (1)$$

Therefore, each mean patch $\overline{P^j}$ will represent a smooth version of their corresponding key-point. The second parameter of the model, that is the mean position $(\bar{x},\bar{y})^j$ of each key-point, is obtained by computing the average of their corresponding coordinates across the n training images.

After creating the model, given a new test image, a correlation score searching scheme is adopted. For each key-point, a set of correlation scores are computed, between their corresponding mean patch $\overline{P^j}$ and the test image in a searching window centred at their mean position $(\bar{x},\bar{y})^j$. Then, a predicted key-point position is given by the coordinates with the highest correlation score. The size of the searching window is a tuning parameter that has to be optimized in the training step.

2.2 Deep learning methods

Regarding the deep learning approaches, two different kinds of deep neural networks were tested, namely a convolutional neural network and a convolutional encoder-decoder network. While CNNs models are used to regress directly the key-points coordinates from the input image, the convolutional encoder-decoder model is used for regression of a density map. Then, the key-points positions are given by local maxima detection on the density map.

2.2.1 CNN model

The implemented convolutional neural network follows the popular CNN architecture for regression, starting from several sequences of convolution-pooling layers to fully connected layers [1]. In this regard, the implemented CNN model is composed by three convolution layers and two

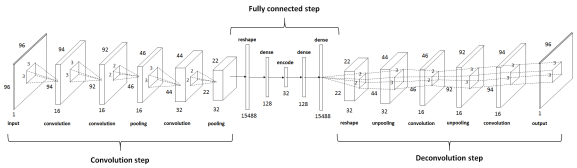


Figure 2: Architecture of the convolutional encoder-decoder model.

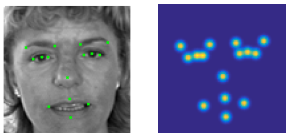


Figure 3: Training stage of the convolutional encoder-decoder model: (left) Training image $I(x)$ with the true key-points coordinates superimposed (green crosses), and (right) The density map $D(x)$, obtained by a superposition of Gaussians at the location of each key-point.

fully connected layers (or dense layers), in which each convolution layer is followed by a 2×2 max-pooling layer. The number of filters is doubled at each convolution layer, ranging from 32, 64 and 128. The size of the filters in each convolution layer is 3×3, 2×2 and 2×2, respectively. Each dense layer has 500 hidden units. The output layer has 30 units, corresponding to the coordinates (x, y) of the 15 key-points.

Attempting to avoid overfitting, we resorted to the utilization of data augmentation and dropout [1]. In this case, all training images along with their corresponding ground-truth key-points coordinates were flipped horizontally, which allows to double the amount of training data.

2.2.2 Convolutional encoder-decoder model

The implemented convolutional encoder-decoder is a deep network that comprises three main steps: i) a convolution step (convolution + pooling), ii) a fully connected encoding step, and iii) a deconvolution step (deconvolution + unpooling). Therefore, a convolutional encoder-decoder can take an image as input and reconstruct another image (with the same size of the input image) as output. The architecture of the implemented autoencoder model is illustrated in Figure 2.

The underlying idea is to use a convolutional encoder-decoder model for regression of a density map (probability density function) of the location of the key-points. This can be viewed as a supervised leaning problem that aims to learn a mapping between an image $I(x)$ and a density map $D(x)$ denoted as $F : I(x) \rightarrow D(x)(I \in \mathbf{R}^{m \times n}, D \in \mathbf{R}^{m \times n})$ for a $m \times n$ pixel image.

For training the model, we create a density map for each training image (see Figure 3). Therefore, for a given training image, each key-point is represented by a Gaussian (with mean at the key-point coordinates and sigma equals 2), and the density map $D(x)$ is formed by the superposition of the Gaussians of each key-point. In the testing stage, the task is to regress this density map from the test image $I'(x)$, and then the key-points locations are given by the local maxima of the density map $D(x)$, as shown in Figure 4.

3 Experimental Results

The dataset used in this work comes from the online Facial Key Point Detection competition of kaggle. The dataset comprises a total of 7049 8-bit grey-scale images.

The parameters of the mean patch searching approach were experimentally tuned using a grid search method, yielding $patch\_size = 10$ and $search\_window = 3$. The implementation of both types of deep nets is based on Lasagne. Nesterov's Accelerated Gradient Descent with momentum is used for optimization, and the mean squared error is used as the objective function to minimize. Regarding the data, the original 8-bit grey-scale images were normalized to the range of $[0, 1]$ and the key-points coordinates to the range of $[-1, 1]$. This data is divided into a training and a validation set, using 20% of the samples for validation.

The experimental results of the implemented facial key-points detection methodologies are presented in Table 1. The results are reported in terms of root mean squared error (RMSE $= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i{}^2)}$, where $\hat{y}$ represents the predicted value and y_i the ground truth).

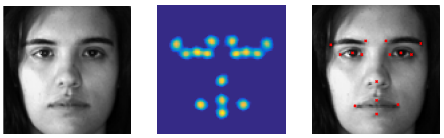


Figure 4: Test stage of the convolutional encoder-decoder model: (left) Test image $I'(x)$, (medium) Estimated density map $D'(x)$, and (right) Key-points prediction by local maxima detection on $D'(x)$ (red crosses).

Table 1: Experimental results of the implemented facial key-points detection approaches. The results are presented in terms of RMSE.

Methodology	Stopping Epoch	Train Loss	Validation Loss	RMSE
Mean Patch Searching	-	-	-	3.79
CNN	5000	0.00108	0.000950	1.48
Convolutional encoder-decoder	4562	0.00162	0.001026	1.66

A first observation is that the mean patch searching algorithm, as expected, is much less robust than both deep learning approaches, yielding by far the highest RMSE (3.79). Regarding the deep learning methodologies, it is possible to observe that the CNN model provided the best results, with a RMSE of 1.48. The convolutional encoder-decoder model provided an RMSE of 1.66. These results demonstrate that the proposed autoencoder model is suitable for facial key-points detection. Although the results are slightly worst than the ones of the traditional CNN model, they are still quite competitive. One reason for this slightly higher RMSE might be explained by the simple local maxima detection algorithm used to detected the key-points positions across the estimated density map. By implementing a more robust technique for this purpose, the obtained results could be better.

4 Conclusions

In this paper, different methods for facial key-points detection were implemented and evaluated, ranging from a classifying search windows method to deep leaning approaches. Our major contribution is the proposal of a convolutional encoder-decoder model for this task. The convolutional encoder-decoder model is used to regress a density map of the key-points location. Afterwards, the facial key-points positions are given by the local maxima of the density map.

Experimental results demonstrate that, although the CNN model achiev a slightly lower RMSE, the proposed convolutional encoder-decoder model has a great potential for facial key-points detection. Contrary to the CNN model, the arquitecture of the convolutional encoder-decoder model does not depend on the number of key-points to be predicted (for instance, the number of units in the CNN output layer have to match the number of key-points). Therefore, the convolutional encoder-decoder model can be used in applications in which the number of key-points, that should be detected, may vary from sample to sample (i.e., when we have to deal with partially occluded facial images).

Acknowledgement

This work was funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF). The first author would like to thank FCT for the financial support of the PhD grant with reference SFRH/BD/102177/2014.

References

[1] Suraj Srinivas, Ravi Kiran Sarvadevabhatla, Konda Reddy Mopuri, Nikita Prabhu, Srinivas Kruthiventi, and Venkatesh Babu Radhakrishnan. A taxonomy of deep convolutional neural nets for computer vision. *Frontiers in Robotics and AI*, 2(36), 2016.

[2] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 3476–3483, 2013.

[3] Michel François Valstar, Brais Martínez, Xavier Binefa, and Maja Pantic. Facial point detection using boosted regression and graph models. In *CVPR*, pages 2729–2736. IEEE, 2010.

[4] Ming-Hsuan Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(1):34–58, Jan 2002.

Epileptic Seizure Prediction with univariated EEG features and Stacked AutoEncoders

Ricardo Barata
rjdb@student.dei.uc.pt
Bernardete Ribeiro
http://www.dei.uc.pt/~bribeiro
António Dourado
https://eden.dei.uc.pt/~dourado/
César Teixeira
https://eden.dei.uc.pt/~cteixei

CISUC – Department of Informatics Engineering
University of Coimbra
Coimbra, PT

Abstract

With this paper we propose an automatic methodology for epileptic seizure prediction based on handcrafted features, extracted from raw electroencephalography (EEG) signals and Stacked Autoencoders. The method consists of manual univariate feature extraction from the raw data, followed by dimensionality reduction, likely allowing the retention of the relevant information and the disregarding of redundant information performed by a two-phase training feed-forward neural network. The method previously mentioned was able to reach statistical significance in 8.3% out of 84 scalp patients and 17.7% out of the 17 invasive patients analysed.

1 Introduction

Despite available drug and surgical treatment options, more than 30% of patients with epilepsy continue to experience seizures [4]. For that matter, if there was a method to warn patients of an impending seizure, or to trigger the administration of an epileptic drug to prevent seizure occurrence, the quality of life of epilepsy patients would benefit of great improvement. Epileptic seizure prediction has been studied from the point of view of computer science and machine learning since the 1970s [5]. Some promising results have been reported, and according to them, a brain state that preceeds seizures was identified and denominated as the pre-ictal state [5]. The pre-ictal state is the target cerebral state in seizure prediction, because once correctly identified it can be said that a seizure has high probability of occurrence on a near future. So far, a lot of experiments have tried to identify the pre-ictal state using machine learning algorithms such as Artificial Neural Network, Support vector machines, among others, based both on raw EEG data and features. A previous study suggests that it is possible to reach sensibilities of 97.5% and false prediction rates of 0.27 h^{-1} [6]. However, the datasets used in this study are very short and non-continuous.

In addition, in 2014 Akara et al. evaluated the use of Stacked Autoencoders for epileptic seizure prediction. Although on their approach the Autoencoders were applied directly on raw data sensitivities between 87% and 100% trading of with high false detection rates were achieved[8]. However, the validation methods of these experiments are not the most conservative, like adjusting parameters on a retrospective way, or the consideration of datasets not large enough to enable trustworthy performance evaluations. The aim of this study is to investigate the success of Stacked Autoencoders on a large scale database, and perform evaluations on an independent validation set, followed by an assessment of the statistical significance.

2 Data and methods

In this section, we describe the data set as well as the hole classification method, from features extraction to evaluation, as described on Figure 1.

2.1 Data set

The database used is part of the EPILEPSIAE database [3] and it is composed by EEG recordings from 101 patients suffering from refractory epilepsy. The recordings encompass 14-121 EEG channels for the patients with intracranial electrodes and 22-37 for the patients with scalp electrodes. For this study, 84 scalp patients and 17 intracranial patients were considered, with an average of 174.35 and 243 hours, respectively, totalling 13.35 million patterns with dimension ranging from 308 to 2662. At least six seizures, per patient, were considered, enabling data partitioning as presented next. The recordings have been previous reviewed by specialized neurophysiologist. Regarding the pre-ictal time or seizure occurrence period (SOP), as it is impossible to be annotated by the neurologists, several values were tested in the range 10 to 60 minutes. From the raw EEG data, 22 univariate features, described on Table 1, were extracted from consecutive 5-seconds windows, without overlapping [9].

Group	Name
Auto-regressive modeling	Predictive error
Decorrelation time	Decorrelation time
Energy	Energy
Hjorth	Mobility
	Complexity
Relative Power	Delta (<4 Hz)
	Theta (4-7 Hz)
	Alpha (8-12 Hz)
	Beta (12-30 Hz)
	Gamma (>30 Hz)
Spectral edge	Frequency
	Power
Statistics	Mean
	Variance
	Skewness
	Kurtosis
Energy of wave coefs.	Decomp. level 1
	Decomp. level 2
	Decomp. level 3
	Decomp. level 4
	Decomp. level 5
	Decomp. level 6

Table 1: Univariative features

2.2 Sparse Autoencoders training

After feature extraction each pattern is a vector with 22 times the number of channel positions, which will be the input of an arrangement of Sparse Autoencoders.

This way the procedure of feature extraction is complemented with an automatic method of dimensionality reduction and relevant information retention. The Sparse Autoencoders are regularized using a L2 regularized term, as well as coefficients of sparsity regularizers. The final network is composed by 3 encoders, with 500, 300 and 100 neuron each, and a Softmax layer. Each autoencoders is trained using a Greedy Layer-Wise Training, on unlabelled data, and the encoder part of each one are stacked together [1]. To this net a Softmax layer, trained with labelled data is added. The hole net is then subject to a process called fine-tuning. The fine-tuning process lightly adjusts the weights of the entire network in order to improve the performance of the binary classification [1], between pre-ictal and non pre-ictal classes.

2.3 Alarm generation

The network will likely misclassify some samples, therefore, to make the classification smoother a sliding window was considered, sample by sample, over the network output. An alarm is raised only if a certain percentage of samples is classified as pre-ictal inside a given window [9]. Also, if an alarm is raised, the next alarm can only take place after a period of time equal to a SOP length after, in order to avoid re-triggering.

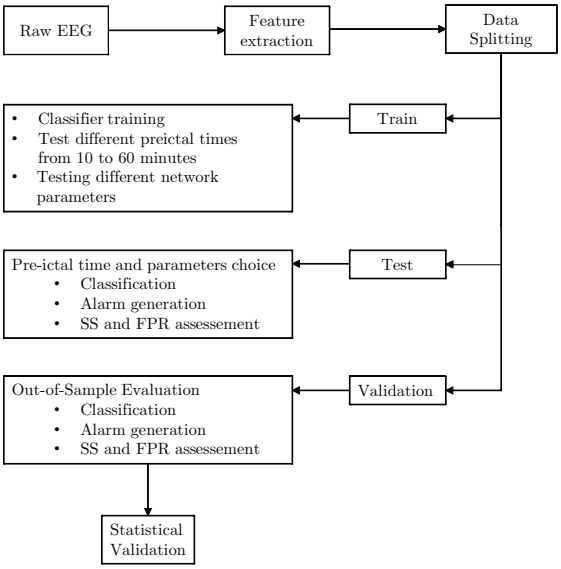


Figure 1: Overall methodology.

2.4 Evaluation

Parameters such as the SOP and the percentage threshold for the alarm generation are selected in the test set, by optimizing two performance measures: sensitivity (SS) and false prediction rate (FPR) [5].

$$SS = \frac{\#PredictedSeizures}{\#TotalSeizures} \times 100 \quad (1)$$

$$FPR = \frac{\#FalseAlarms}{InterictalTime - (\#Seizures \times SOP)} \quad (2)$$

These parameters are then applied along with the classifier to perform the classification on the validation test. Again sensitivity and FPR are assessed, and the results are statistically validated using an analytic random predictor based on a binomial distribution [2, 7].

3 Results and Discussion

The validation data set consisted of 4362.78 hours of interictal data and 376 seizures, of which 99 (26.3%) were successfully predicted, for the scalp EEG patients and 1570.24 hours of interictal data and 148 seizures, of which 30 (20.3%) were successfully predicted, for the intracranial EEG patients, as presented on Table 2.

Regarding the patients with scalp recordings the median sensibility and false prediction rate accomplished were 16.67% and 0.20 h^{-1}. Regarding patients with intracranial records the median sensibility and false prediction rate were 8.33% and 0.15 h^{-1}, respectively.

	Scalp	Intracranial
Patients	84	17
Median Sensibility	16.17%	8.33%
Median FPR	0.20	0.15
Interictal period (hours)	4362.78	1570.24
Seizures in validation	376	148
Predicted seizures	99	30
Patients with significant results	7	3

Table 2: Results summary.

Comparing with the random predictor the method was able to identify pre-ictal periods above chance level in 3 out of 17 patients with intracranial EEG recordings (17.65%) and 7 out of 84 patients with scalp EEG recordings (8.3%). These results can be considered significant according to the binomial test at the 5% significance level (p-value$\leq$0.05). In the light of these results, we can assume that for some patients it is feasibly to perform seizure prediction using the method described, which is based on a feature extraction and deep artificial neural networks. It is also important to refer that sensitivities of 97.5% and FPR of 0.27, so far reported by other authors, h^{-1} does not carrier the same weight as the one here proposed. The presented study was performed on a large database with a large number of patients and long-term recordings. Also data used here was not biased regarding the existence inter-ictal periods. We have also statistically evaluated the performance of the classification on independently data set, according to a scheme proposed by Schelter et al. (2006) using a random predictor.

Another advantage of our approach was the automatic electrode and feature selection. Extract this number of features from a complete channel record leads to a considerably large dimensionality, which usually disables the use of classical classification methods. However, by making use of this advanced system of dimensionality reduction and at the same time resorting to powerful GPUs we were able to train all the classifiers needed in reasonable timing.

4 Conclusion

From the very beginning epileptic seizure prediction has proven to be a very complicated and heavy task. Throughout the years developments have been made trying to improve the performances of automatic seizure prediction, with the intention of achieving a system reliable enough to be used on a daily basis by patients suffering from seizures. The results presented on this study prove that, to a certain point, it is possible to predict seizures on some patients if the proper methodologies and validation systems were used. The numbers may not be that impressive, but if 8% of the patients suffering from refractory epileptic seizures cloud have access to a system able to give feedback about their condition they could be assisted quickly and perhaps even before the seizure took place.

References

[1] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle, et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19:153, 2007.

[2] Hinnerk Feldwisch-Drentrup, Björn Schelter, Michael Jachan, Jakob Nawrath, Jens Timmer, and Andreas Schulze-Bonhage. Joining the benefits: combining epileptic seizure prediction methods. *Epilepsia*, 51(8):1598–1606, 2010.

[3] Juliane Klatt, Hinnerk Feldwisch-Drentrup, Matthias Ihle, Vincent Navarro, Markus Neufang, Cesar Teixeira, Claude Adam, Mario Valderrama, Catalina Alvarado-Rojas, Adrien Witon, et al. The epilepsiae database: an extensive electroencephalography database of epilepsy patients. *Epilepsia*, 53(9):1669–1676, 2012.

[4] Patrick Kwan and Martin J. Brodie. Early identification of refractory epilepsy. *New England Journal of Medicine*, 342(5):314–319, 2000.

[5] Florian Mormann, Ralph G. Andrzejak, Christian E. Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130 (2):314–333, 2007.

[6] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770, 2011.

[7] Björn Schelter, Matthias Winterhalder, Thomas Maiwald, Armin Brandt, Ariane Schad, Andreas Schulze-Bonhage, and Jens Timmer. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1):013108, 2006.

[8] Akara Supratak, Ling Li, and Yike Guo. Feature extraction with stacked autoencoders for epileptic seizure detection. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4184–4187. IEEE, 2014.

[9] César Alexandre Teixeira, Bruno Direito, Mojtaba Bandarabadi, Michel Le Van Quyen, Mario Valderrama, Bjoern Schelter, Andreas Schulze-Bonhage, Vincent Navarro, Francisco Sales, and António Dourado. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Computer methods and programs in biomedicine*, 114(3):324–336, 2014.

Boosting Compression-based Classifiers for Authorship Attribution

Filipe Teixeira
fmteixeira@ua.pt
Armando J. Pinho
ap@ua.pt

IEETA/DETI
University of Aveiro
Aveiro, Portugal

Abstract

Authorship attribution is the task of assigning an author to an anonymous document. Although the task was traditionally performed by expert linguists, many new techniques have been suggested since the appearance of computers, in the middle of the 20^{th} century, some of them using compressors to find repeating patterns in the data. This work will present the results that can be achieved by a collaboration of more than one compressor using a meta-algorithm known as Boosting.

1 Introduction

The task of authorship attribution, i.e. assigning an author to an anonymous document, can be approached in many different ways. In this work, compression-based classifiers were used. These classifiers use general purpose data compressors to measure the similarity between a pair of documents, which allows them to use a set of reference documents, whose authorship is undisputed, to assign the most likely to an anonymous text, provided that the real author has texts in the reference set. Note that if the real author has no documents in the reference set, one is still selected.

2 Dataset

Tests were performed using Varela's dataset [9]. The dataset has a total of 3000 texts by 100 authors, and each author writes about one of ten themes: Economy, Gastronomy, Health, Law, Literature, Politics, Sports, Technology, Tourism and Unspecified Subject. Some relevant statistics about the dataset are in Table 1.

	Bytes	Tokens
Mean	3000	593
Standard Deviation	1536	299
Median	2806	556
Minimum	209	39
Maximum	17866	3470
Total	9000755	1780397

Table 1: Varela's dataset descriptive statistics.

Tests performed in this work used seven random references by each author, and the remaining 23 as testing targets, for a total of 700 references and 2300 targets.

3 Classifiers

The classifiers used in this work have five components, two for preprocessing and three for classification. Before classifying the target documents, the classifier can concatenate all the reference texts by the same author, resulting in one text by author, and normalize the texts, assuring that every author has the same amount of reference data. The normalization process is different depending on the concatenation choice. If the references are not concatenated, each one is truncated to have the same size as the smallest. If they are concatenated, each reference is truncated at a length proportional to it's size before being concatenated, otherwise entire references could be eliminated from the process. The two main steps in the classification process require the remaining three components: a data compressors and a similarity measure to rank the references by similarity, and an evaluation method that uses such rank to select the most likely author.

3.1 Similarity Measures

In this work four different measurements of the similarity between x and y were used. Normalized Compression Distance (1) [1] is defined as,

$$NCD(x,y) = \frac{C(x\cdot y) - \min\{C(x),C(y)\}}{\max\{C(x),C(y)\}}, \quad (1)$$

where $C(x)$ is the length of x after being compressed by some compressor C and $x\cdot y$ is the concatenation of y to x, i.e. x followed by y, without any connecting characters. The following measurement is known as Conditional Complexity of Compression (2) [3],

$$CCC(x,y) = C(y\cdot x) - C(y). \quad (2)$$

Another measurement, Normalized Conditional Compression Distance (3) [8] , is defined as

$$NCCD(x,y) = \frac{\max\{C(x|y),C(y|x)\}}{\max\{C(x),C(y)\}}, \quad (3)$$

where $C(x|y)$ is the length of x when compressed using both x and y's models. The last measurement is the Normalized Relative Compression (4) [7]

$$NRC(x,y) = \frac{C(x||y)}{|x|}, \quad (4)$$

where $C(x||y)$ is the exclusive conditional compression, and represents the length of x after being compressed using only y's models.

3.2 Compressors

Five compressors were tested: gzip, bzip2, LZMA, PPMd and CondCompNC, all with the maximum compression setting available, except for CondCompNC which was used with the options presented in Table 2, shown to facilitate replicability.

Mode	Options
$C(x\cdot y)$	-tt 1:2 -tt 1:3 1/10 -tt 1:4 1/100
$C(x\|y)$	-rt 1:2 -rt 1:3 1/10 -rt 1:4 1/100 -tt 1:2 -tt 1:3 1/10 -tt 1:4 1/100
$C(x\|\|y)$	-rt 1:2 -rt 1:3 1/10 -rt 1:4 1/100

Table 2: Options used with CondCompNC.

These compressors were selected to experiment with different compression methodologies and, in the case of CondCompNC, because it offers features required by some similarity measures.

3.3 Evaluation Methods

For the last element, evaluation methods, three variations were tested. The first, called Maximum Similarity, selects the author by looking only at the most similar reference. The second, Equal Voting, selects the N most similar references and uses a majority voting to elect the author. The last method is called Author's Average and also uses the N most similar references. Those references are then grouped by author and the group with best average similarity is used to assign the author.

4 Committees

Looking at existing work [6], including our own, it's clear that even the strongest classifiers misclassify target documents that weaker classifiers were successful at. By taking into account the classifications of more than one classifier it may be possible to improve the performance achieved by any single classifier. Using the classifiers presented before, 96 different

combinations can be made, ignoring different values of N in some evaluation methods. After running all the classifiers, an upper bound for the committee can be set by checking if any of the classifiers correctly classified each target document. The results are in Table 3, where C_1, C_2, C_3 represent the three best classifiers found after testing all 96 combinations.

Classifier	Performance
C_1	68.87%
C_2	68.39%
C_2	68.04%
Disjunction	94.86%

Table 3: Upper-bound established by the disjunction of all classifiers.

4.1 Boosting and AdaBoost.M2

Boosting is a meta-algorithm that deals with a relevant problem, using an ensemble of weak classifiers to build a strong classifier. The algorithm requires a set of training documents, used to compute the weights assigned to each classifier. Knowing the weights, classifying a target document starts by running each classifier independently and then merging their answers using the weights and some merging strategy.

AdaBoost.M2 [2] is a Boosting algorithm, and it can deal with multiclass problems, which is required in this work. One of the steps in the algorithm requires classifiers to provide plausibility measure, i.e. given a target document and the set of possible authors, assign a value to each author, representing how plausible he is. This plausibility is not a probability, it's only a mechanism to allow a classifier to share more information with the boosting process. As such, the base classifiers need to be modified, to return a plausibility vector instead of the most likely author. In this work, this was achieved by normalizing the distances to the interval $[0, 1]$, where zero is the least plausible and one the most. The algorithm iterates T times, on each iteration one classifier's final weight is learned and that classifier is then removed from the pool. Using $T = 96$, all available classifiers, and the references set used for training, with 300 training references and 400 training targets, the weights learned are shown in Figure 1. In AdaBoost.M2, lower weights represent a better classifier.

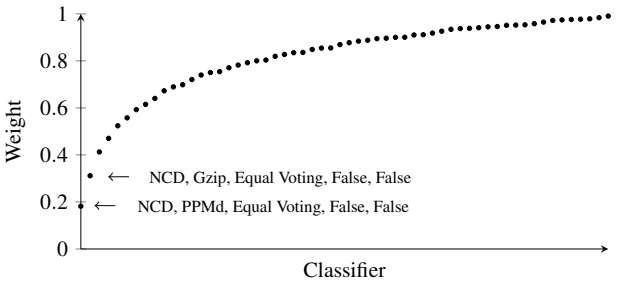


Figure 1: Weights assigned to classifiers by AdaBoost.M2.

In Figure 2 we see that, overall, better classifiers tended to be assigned lower weights, as expected.

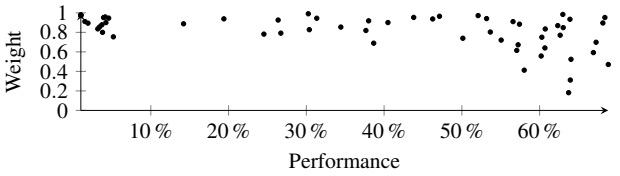


Figure 2: Performances achieved individually vs. weight assigned.

4.2 Results

With the weights presented in the previous section, the committee was able to correctly classify 76.29% of the tests, 7% above the best classifier. Although [5, 6] obtained better results for single classifiers, using the same methodology and dataset, we couldn't replicate them and as such the comparison is made to ours.

Category	Commiitee	Max.	Voting	Average
Economy	75.11%	69.87%	62.01%	65.50%
Gastronomy	61.14%	61.57%	41.05%	49.78%
Health	80.00%	69.57%	65.65%	65.65%
Law	74.12%	72.81%	64.47%	64.03%
Literature	62.28%	56.58%	47.81%	51.31%
Politics	78.26%	66.09%	66.52%	74.78%
Sports	81.74%	73.48%	76.52%	75.65%
Technology	84.78%	80.87%	79.13%	76.95%
Tourism	82.17%	72.61%	73.04%	70.00%
Unspecified Subject	83.04%	65.22%	76.09%	78.69%
Average	76.29%	68.87%	65.26%	67.26%

Table 4: Results by theme from a committee with 96 classifiers mixed by *AdaBoost.M2*

Table 4 shows the results achieved by the committee learned with AdaBoost.M2, and the best classifier from each evaluation method, with the targets grouped by theme. We see that, with the exception of Gastronomy, the committee provided better results in every theme. For clarification, the targets are still classified by the author, not the theme. It's displayed by theme due to the large number of authors. Other variants were tried, limiting the number of classifiers in the committee, but the results were similar, as can be seen in Table 5.

Number of classifiers	Performance
All (96)	76.29%
16	76.59%
10	73.54%

Table 5: Performance achieved by using a different number of classifiers in the committee.

Although the results were similar, reducing the number of classifiers greatly reduces the computation time required to classify a document, while providing similar accuracy. The results may be improved by using other pruning methods [4], however this was not tested.

5 Acknowledgments

This work was partially funded by the FCT - Foundation for Science and Technology, in the context of the projects UID/CEC/00127/2013 and PTDC/EEI-SII/6608/2014.

References

[1] Rudi Cilibrasi and Paul MB Vitányi. Clustering by compression. *IEEE Transactions on Information theory*, 51(4):1523–1545, 2005.

[2] Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning (ICML 1996)*, pages 148–156, 1996.

[3] Mikhail B Malyutov, Chammi I Wickramasinghe, and Sufeng Li. Conditional complexity of compression for authorship attribution. 2007.

[4] Dragos D Margineantu and Thomas G Dietterich. Pruning adaptive boosting. In *ICML*, volume 97, pages 211–218, 1997.

[5] W Oliveira, Edson Justino, and Luiz S Oliveira. Comparing compression models for authorship attribution. *Forensic science international*, 228(1):100–104, 2013.

[6] W Oliveira Jr, E Justino, and L Oliveira. Authorship attribution of documents using data compression as a classifier. In *Proceedings of the World Congress on Engineering and Computer Science*, volume 1, 2012.

[7] Armando J. Pinho, Diogo Pratas, and Paulo JSG Ferreira. Authorship attribution using relative compression. In *Proceedings of the Data Compression Conference, 329-338*, pages 329–338, 2016.

[8] Diogo Pratas and Armando J Pinho. A conditional compression distance that unveils insights of the genomic evolution. In *Proceedings of the Data Compression Conference*, page 421, 2014.

[9] Paulo Júnior Varela. O uso de atributos estilométricos na identificação da autoria de textos, 2010.

Detection of small juxta-pleural nodules in computed tomography images

Guilherme Aresta[1,3]
bio11017@fe.up.pt

António Cunha[2,3]
acunha@utad.pt

Aurélio Campilho[1,3]
campilho@fe.up.pt

[1]Faculdade de Engenharia da Universidade do Porto
Porto, Portugal

[2]Universidade de Trás-os-Montes e Alto Douro
Vila Real, Portugal

[3]INESC-TEC - INESC Tecnologia e Ciência
Porto, Portugal

Abstract

A method for the detection of small juxta-pleural nodules is proposed. The method is developed and tested using the public Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset, by creating a sub-dataset of juxta-pleural nodules. The lung volume is segmented using region-growing and refined with morphological operations and active contours to include juxta-pleural nodules. Nodule candidates are searched slice-wise inside the lung volume segmentation. Solid nodules are detected by selecting an appropriate threshold inside a representative sliding window. Sub-solid and non-solid nodules are enhanced with a multiscale Laplacian-of-Gaussian filtering prior to their detection. Obvious non-nodule candidates, namely corresponding to small blood vessels, are discarded using fixed rules. Then, a support vector machine with radial basis function is trained with the remaining candidates to further reduce the number of false positives (FPs). The majority of the studied juxta-pleural nodules have solid texture. The initial candidate detection step achieves a sensitivity of 92% with 3450±2720 FPs/scan. Fixed rules reduction drops the sensitivity to 72.5% with 95.5±52.1 FPs/scan. The final system sensitivity is 57.4% with 4 FPs/scan, with an average sensitivity score of 0.39. The performance is similar or better than state-of-the-art methods, especially when considering the high number and small radius of the studied juxta-pleural nodules.

1 Introduction

Lung cancer is the most lethal type of cancer [9], demanding for early detection in order to improve the survival rate. However, factors such as the limitations of the human visual system hinder the nodule detection by physicians, motivating the need for computer-aided detection (CADe) systems. Lung nodules can be classified according to their radius, texture and location. In terms of radius, r, nodules can be large ($r > 5mm$) or small ($r \leq 5mm$). Large solid nodules tend to be simpler to detect [8]. Texture-wise, nodules are solid, sub-solid or non-solid [1]. Solid nodules have well defined boundaries and good contrast with the lung parenchyma. Sub-solid and non-solid nodules have blurry limits and are less contrasted with the parenchyma, thus being more difficult to detect. Location-wise, nodules are juxta-vascular when attached to blood vessels, peri-fissural when they are near fissures, juxta-pleural when they are attached to the pleura or, by exclusion, isolated [10]. From these, juxta-pleural are the most difficult to detect due to their peripheral location, intensity similarity with non-parenchymal tissue and shape variety. The combination of CADe systems tends to outperform isolated systems [10]. In order to contribute for this combined multi-detector system, this paper focuses on the detection of small juxta-pleural nodules.

2 Materials and Methods

A dedicated method for the detection of small juxta-pleural nodules is presented. The method is developed and evaluated using the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset. The LIDC-IDRI dataset [1] has 1012 cases and over 2500 nodules with $r > 1.5mm$ segmented by 4 specialists. The nodules are subjectively graded on several properties, including texture and subtlety. No position-based classification is provided and thus a subset of juxta-pleural nodules is proposed. First the lung volume is segmented. Then, solid nodules are detected via direct threshold while sub-solid and non-solid nodules are enhanced with a multi-scale Laplacian-of-Gaussian filtering. Finally, the resulting false positives (FPs) are reduced with fixed rules and a support vector machine.

2.1 Lung volume segmentation

The initial lung volume segmentation is performed as in [5] and then refined slice-wise using Chan-Vese active contours [2]. First, a seed point in the non-parenchymal lung tissue is automatically selected for intensity-based region-growing. The resulting binary mask is inverted to obtain the segmentation of the lung volume. Cases where the two lungs are connected are solved by iteratively eroding the segmentation until two connected components are obtained. Then, a parenchymal region-growing is applied to obtain the original segmentation. Juxta-pleural nodules are included by performing a 2D morphological closing operation with a large diameter circle. The segmentation is refined using the Chan-Vese active contour to compensate local imperfections caused by the closing.

2.2 Dataset of juxta-pleural nodules

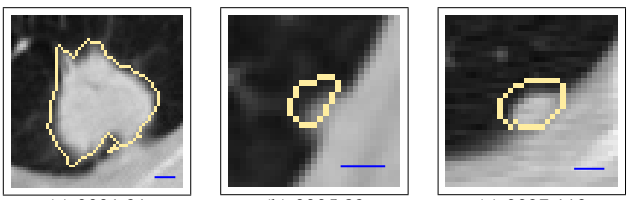
We consider that a lung nodule is juxta-pleural if the distance from the pleural wall is $< 1.5mm$. For that purpose, the boundary of the lung volume segmentation is dilated with disk of radius $1.5mm$, corresponding to the smallest nodule of the dataset. All ground-truth nodules with at least 1 voxel intersecting the dilation are considered juxta-pleural. The dataset is then manually revised to exclude all nodules that never contact with the pleura. Fig. 1 shows examples of juxta-pleural nodules.

2.3 Candidate detection

All nodule candidates are searched inside the lung volume segmentation. Solid candidates are searched slice-wise inside a fixed width sliding window with stride 1. The contrast inside the square region is saturated at 1% of the minimum and maximum values and the candidate is then detected by selecting an appropriate threshold using the Otsu's method [6]. The segmentation of all windows is combined using the OR logical operation. Sub-solid and non-solid nodules are enhanced with a multi-scale Laplacian-of-Gaussian (LoG) filtering. Filters are not normalized by the kernel's standard deviation prior to the enhancement, heavily prioritizing small structures. The maximum response of all filters is combined. Then, the filtering and the selection of an appropriate threshold value using Otsu's method are performed slice-wise. The lists of the 3D connected components corresponding to solid and sub-solid/non-solid nodules are combined. A detection example is shown in Fig. 2.

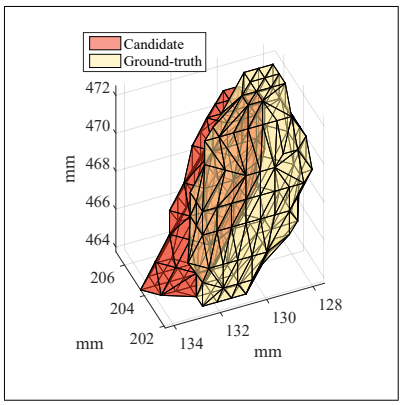
2.4 False positive reduction

The proposed candidate detection method produces a high number of false positives, which correspond mainly to small blood vessels along the lung parenchyma. The most obvious non-nodules are removed with fixed rules. Candidates with equivalent $r > 6mm$ and with a distance greater than $6mm$

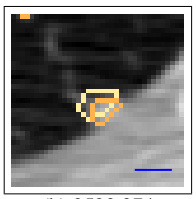


(a) 0001,91 (b) 0005,89 (c) 0027,113

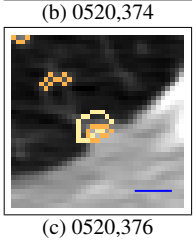
Figure 1: Examples of nodules considered as juxta-pleural. Each example is retrieved from a scan and slice (LIDC-IDRI case#, slice#). Blue scale bar corresponds to $5mm$.



(a) 3D reconstruction



(b) 0520,374



(c) 0520,376

Figure 2: 3D reconstruction of a juxta-pleural lung nodule candidate (orange) along with the ground-truth (yellow). The correspond initial, middle and final slices are shown. Each example is retrieved from a different slice (LIDC-IDRI case#, slice#). Blue scale bar corresponds to 5*mm*.

to the pleura wall are removed. Likewise, all candidates with $r < 1.5mm$, corresponding to the smallest nodule in the LIDC-IDRI, are removed. A support-vector machine with radial-basis-function kernel (SVM-rbf) is trained with features extracted from the remaining candidates. The features used to train the classifier are related to intensity, Hessian values, intensity gradient, geometry and distance to the pleura.

3 Results

The method is evaluated on 729 scans of the LIDC-IDRI dataset. The remaining 283 scans were not used due to file reading problems. From these, 315 scans contain juxta-pleural nodules with $r \leq 5mm$, corresponding to a total of 510 nodules. Approximately 80% of the nodules have solid texture and the remaining 20% are equally divided in sub-solid and non-solid. Juxta-pleura nodules show lower average inter-observer agreement in comparison to all nodules of the dataset. Solid candidates are searched inside a $30 \times 30mm$ sliding window. For sub-solid and non-solid nodules the σ values of the LoG filters are set to $\{1, 1.5, 2\}$ voxels. The initial candidate detection method has an overall sensitivity of 92%. The majority of the failed nodules are non-solid. Candidates are considered detected (true positive, TP) if at least 1 voxel intersects with the ground-truth. The average segmentation Dice coefficient is 0.22 ± 0.14. Visual analysis shows that solid nodules are properly segmented and that the Dice coefficient is affected by displacements of the ground-truth. In the first stage, an average of 3450 ± 2720 FPs/scan is detected. After the fixed rule reduction this value drops to 95.5 ± 52.1 FPs/scan. The corresponding system sensitivity is 72.5%. The classifier is trained using 2-fold cross validation scan-wise. The procedure is repeated $50\times$ with random sets. The classifier has an AUC of 0.95 ± 0.01, similar or better than other state-of-the-art methods [8-10]. The supervised learning FP reduction allows to achieve an overall sensitivity of 57.4% with 4 FPs/scan and 61.8% with 4 FPs/scan if only solid nodules are considered.

The performance of the system is evaluated in terms of sensitivity vs FPs and using a score metric, the average sensitivity at $2^{-3\ldots3}$ FPs/scan [10]. The results in Table 1 show that the performance of our method is similar or better than others, especially considering the large number of scans and the small radius of the nodules studied. Methods **C** and **H** mainly consider solid nodules with $r \geq 5mm$, increasing the methods' overall performance since these nodules are easier to detect [8]. Method **H** does not use the LIDC-IDRI dataset and thus the results are not directly comparable. Despite that, we achieve a score of 0.42 if only solid nodules are considered, which is similar to their result. Furthermore, method **C** considers juxta-pleurals nodules as any abnormality inside a 5-layer erosion of the lung volume. Consequently, their dataset includes nodules that never contact with the pleural wall. The higher nodule radius and the differences in the dataset contribute to the differences between the results.

Table 1: Juxta-pleural lung nodule detection performance of different systems. r is the nodule radius. Nod. is the total number of studied nodules. Sens. is the sensitivity (%). Rad. is the nodule radius, in *mm* A - [7], B - [3], C - [4], D - Fujitalab, E - Region growing volume plateau, F - Channeler Ant model, G - Voxel-based neural approach, H - ISI-CAD, I - Philips Lung Nodule CAD ([10]). A and B used a private dataset. C used the LIDC-IDRI. The remaining systems used the ANODE09 dataset.

	# Scans	Nod.	Rad.	FPs/scan	Sens.	Score
A	42	25	≥ 2.5	6	72.0	-
B	-	-	-	-	66.5	-
C	**205**	**323**	≥ 1.5	**4.1**	**89.2**	**-**
D	50	60	≥ 4	4	15.3	0.10
E	50	60	≥ 4	4	33.9	0.16
F	50	60	≥ 4	4	35.6	0.21
G	50	60	≥ 4	4	35.6	0.19
H	**50**	**60**	≥ 4	**4**	**69.5**	**0.44**
I	50	60	≥ 4	4	22.9	0.14
Ours	**315**	**510**	[1.5-5]	**4**	**57.4**	**0.39**

4 Conclusion

A method for the detection of small juxta-pleural nodules in CT images is presented. Nodule candidates are searched slice-wise by selecting an appropriate threshold inside a sliding window. The more subtle nodules are enhanced with LoG filtering prior to detection. A proper lung volume segmentation and dedicated FP reduction steps allow the system to achieve similar or better performance than the state-of-the-art. The method can be further improved by refining the lung volume segmentation to include nodules located on the vertexes of lungs and by improving the segmentation of the candidates to increase the sensitivity after FP reduction.

Acknowledgments

This work is financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-016673.

References

[1] S. G. Armato, G. McLennan, and L. Bidaut 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(2):915, 2011.

[2] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[3] G. De Nunzio, A. Massafra, R. Cataldo, and et al. Approaches to juxta-pleural nodule detection in CT images within the MAGIC-5 Collaboration. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648:S103–S106, 2011.

[4] H. Han, L. Li, and F. Han et al. Fast and Adaptive Detection of Pulmonary Nodules in Thoracic CT Images Using a Hierarchical Vector Quantization Scheme. *IEEE Journal of Biomedical and Health Informatics*, 19(2):648–659, 2015.

[5] J. Novo, J. Rouco, A. Mendonça, and A. Campilho. Reliable lung segmentation methodology by including juxtapleural nodules. *Springer International Publishing, Lecture Notes in Computer Science*, 8815:227–235, 2014.

[6] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[7] A. Retico, P. Delogu, and M.E. Fantacci et al. Lung nodule detection in low-dose and thin-slice computed tomography. *Computers in Biology and Medicine*, 38(4):525–534, 2008.

[8] A. Setio, C. Jacobs, and J. Gelderblom et al. Automatic detection of large pulmonary solid nodules in thoracic CT images. *Medical Physics*, 42(10): 5642–5653, 2015.

[9] R. L. Siegel, K. Miller, and A. Jemal. Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians*, 65(1):5–29, 2015.

[10] B. van Ginneken, S. G. Armato, and B. de Hoop et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: The ANODE09 study. *Medical Image Analysis*, 14(6):707–722, 2010.

Deriving ECG to compute inhalation during fire experiments

Raquel Sebastião
raquel.sebastiao@ua.pt

Institute of Electronics and Informatics Engineering of Aveiro (IEETA)
Department of Electronics, Telecommunications and Informatics (DETI)
University of Aveiro,
3810-193 Aveiro, Portugal

Sandra Sorte
ssss@ua.pt

Joana Valente
joanavalente@ua.pt

Ana I. Miranda
miranda@ua.pt

Centro de Estudos do Ambiente e do Mar (CESAM)
Department of Environment and Planning (DAO)
University of Aveiro,
3810-193 Aveiro, Portugal

José Maria Fernandes
jfernan@ua.pt

DETI / IEETA - University of Aveiro

Abstract

When fighting forest fires, firefighters are exposed to several pollutants at different concentrations, which can raise serious health problems. The objective of this work is to show that is possible to estimate the firefighters' pollutants inhalation when in forest fire fights based on online monitoring of environmental (CO) and physiological information (ECG). Therefore, the firefighters were monitored during experimental forest fire fights. Through the physiological data, ECG-derived respiration (EDR) was estimated based on the area under the QRS complex. Thereafter, when associated with exposure to CO (carbon monoxide), the smoke inhalation was computed. The analysis of smoke inhalations revealed meaningful insights: it is possible to detect extensive exposures and to identify critical risks of faint due to smoke inhalation and hazardous conditions. Moreover, the results discovered lead to the conclusions that the continuous monitoring of such information will help to make a safe and successful management of the firefighters and of the forest fire.

1 Introduction

During forest fire fights, firefighters (FF) are exposed to high levels of pollutants, which can raise harmful healthy problems [7],[8],[12]. Recent research points to that high smoke exposure decrease respiratory capability [2],[6],[11]. Therefore, monitoring pollutants exposure can be useful to ensure high exposure to inhaled pollutants is quickly identified to avoid dangerous situations.

In this work, by combining of physiological and environmental data, we show that is possible to perform online estimation of the inhalation of CO by firefighters when involved in firefighting.

In our scenario, a team of 4 male firefighters (FF) was monitored during an experimental field burnings of Gestosa 2015 [8], [9], in central Portugal. The 4 FF have ages ranging from 19 to 36 years old (24±8) and height from 170 to 191 cm (175±10). Each FF was equipped with the VR2 system [5] for the collection of several data (GPS data, environmental measurements and physiological signals - ECG , HR and body temperature). The CO concentration was monitored using the GasAlertExtreme CO equipment from BW technologies. Carbon monoxide is present in all fire environments and the knowledge of CO peak concentration values to which firefighters are exposed is of major importance, given the risk of asphyxia.

The evaluation of the exposure of firefighters to smoke requires the comparison with occupational exposure standard (OES) values defined for air pollutants. According to the American Conference of Governmental Industrial Hygienists (ACGHI), OES are presented as the: (i) threshold limit value (TLV) for a 8-hours' time-weighted average (TWA); (ii) TLV of a 15 minutes short-term exposure limit (STEL); and (iii) peak limit. These guidance values are established aiming at human health protection and, for CO, the values are presented in Table 1.

Table 1 - Occupational exposure standard (OES) limits for the CO.

CO pollutant	
TLV-TWA (ppm)	25
TLV-STEL (ppm)	200
Peak Limit	400

Respiratory and Pollutants Inhalation Estimation

As measuring respiratory information during forest fire fights is impracticable, we estimated the respiratory signals using information derived from ECG. ECG-Derived Respiration (EDR) methods exploit the respiratory induced changes of the ECG to provide the estimation of the respiratory rate and the temporal pattern of the respiration [10],[3]. The Matlab® code used to compute the EDR, through a method based on the QRS area [10] is available from the Pshysionet [4] at [1].

From EDR it is possible to extract the respiratory rate (RR). Thereafter, based on average respiratory volume for normal healthy individuals we were able to estimate the respiratory minute ventilation (RMV) for each FF. The CO inhalation is computed by multiplying the CO concentration by RMV along time intervals. More details can be found in [13]. The relation between heart rate (HR) and RR is shown in Figure 1 for FF3, where smaller HR values are related with lower number of breathings.

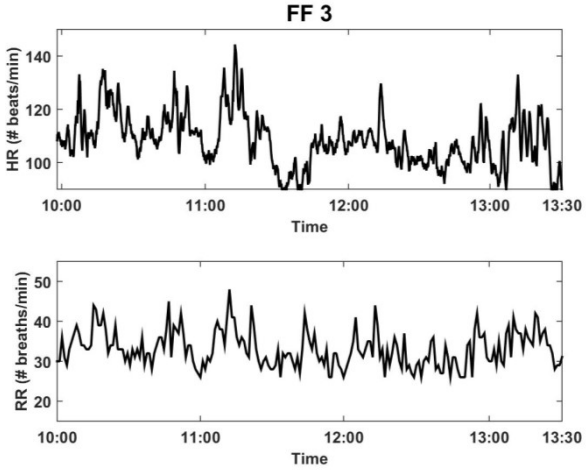


Figure 1 - Heart Rate (HR) and Respiratory Rate (RR) for FF3.

2 Results and Discussion

Synchronizing the raw information, it was possible to obtain a multimodal dataset for each FF with both ECG heart rate variations and the estimated CO exposure profile.

We present the profiles for firefighter FF2 and FF3. FF2 was located on a low exposure position as backup and he was not directly attacking the fire during the trial. FF3 was close to the active fire and smoke and he was ready to attack the fire if needed. The CO concentration (ppm) and inhalation CO are presented in Figure 2 (a) for FF2 and in Figure 2 (b) for FF3. The CO exposure values are much smaller (one order of magnitude lower) than the ones affecting FF3: the FF2 values are mostly bellow the TWV-TLA level, while most of the FF3 values are well above it.

Thus the focus was on FF3. The exposure values monitored for FF3 are very high, with peak values reaching 200 and 300 ppm, which indicates an inhalation risk situation. This firefighter was located in a position influenced by the smoke plume, since the wind was blowing from South, exposing the firefighters that were positioned upwind in the experimental parcel. As show in Figure 2 (b), FF3 was exposed to CO concentrations up to 355 ppm. This peak concentration values can be of major importance, in particular for identifying the risk of asphyxia. Although these higher values occurred only four times and for very short periods of time, a longer exposure to such high CO concentrations suggests a health risk situation. Therefore, controlling such exposures is of utmost relevance for the firefighters' health.

Wind direction and the distance to the fire play an important role in the CO inhalation. As closer to the fire, the higher the inhalation. At the same time, if the firefighter is in a position buffeted by the smoke, the pollutants inhalation will increase. Attempting to reduce the smoke exposure and the risk associated with the inhalation of pollutants, a decision system may be a viable option. With such a system, several signals are continuously monitored and analysed in order to detect events or changes that may be correlated with health risks. Thereafter, in such cases, an alarm can be trigger and the substitution of the firefighter can be done more promptly. The results achieved with this work contribute to the development of an advisory system to diminishing the health risks which firefighters are exposed and improve their well-being in firefighting scenarios.

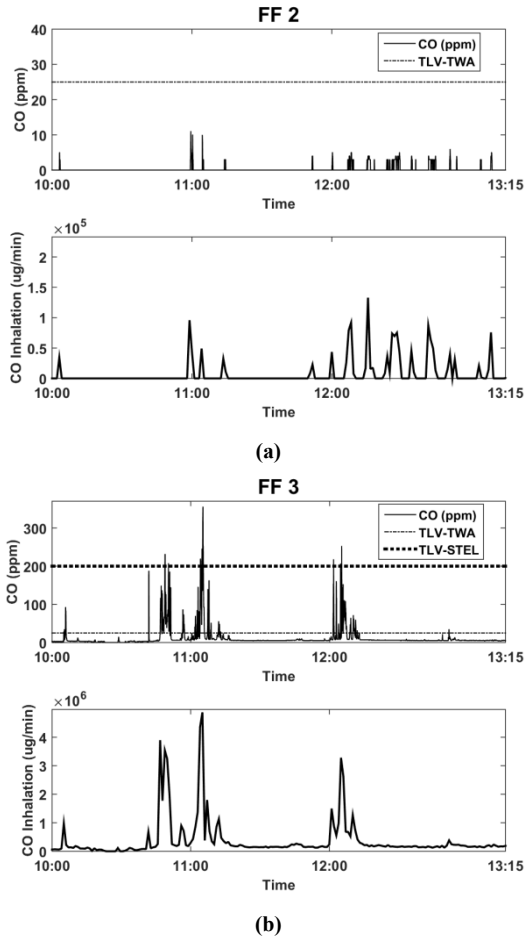


Figure 2 - CO concentration (ppm) and inhalation for firefighter FF2 (a) and FF3 (b). FF2 was not exposed to smoke and for FF3, exposed to smoke close to fire. Note that the graphics in (a) and (b) have different ranges.

3 Conclusions

The dynamics behind the data collected are challenging. The analysis of ECG based and derived signals (EDR) can be paramount importance for performing online detection of critical changes, namely those related with physiological and exposure conditions of FF.

The obtained results show that (even in trial conditions) FF can find situations that are of utmost interest to monitor, namely smoke inhalation in order to avoid adverse health side-effects. In operational scenarios, the exposure to several pollutants and to different concentrations of these, can raise hazardous health problems to firefighters. Performing online monitoring may allow prompt detection of critical situations, specially those where extensive exposures can or are occurring and avoid actively situations like fainting due to smoke intoxication.

With the technical monitoring solution in place, the next research step will be devoted to the development of a decision support system to be applied in real-time during firefighting scenarios helping to make a safe and successful management of the firefighters and of the forest fire.

4 Acknowledgments

A particular thank is due to Domingos Xavier Viegas and his research team for the organization and for performing the Gestosa experiments.

This work was supported by the Portuguese Science Foundation (FCT) through national funds, and co-funded by the FEDER, within the PT2020 Partnership Agreement and Compete 2020 under projects IEETA (UID/CEC/00127/2013), CESAM (UID/AMB/50017/2013), VitalResponder2 (PTDC/EEI-ELC/2760/2012), VR2market (CMU Portugal program, CMUP-ERI/FIA/0031/2013). The Post-Doc grants of R. Sebastião and J. Valente (BPD/UI62/6777/2015 and SFRH/BPD/78933/2011, respectively) are also acknowledged.

References

[1] ECG-derived respiration. 2015. Retrieved October 20, 2015 from https://www.physionet.org/physiotools/edr/

[2] C.E. Bergstrom, A. Eklund, M. Skold, and G. Tornling. 1997. Bronchoalveolar lavage findings in firefighters. *Am J Ind Med* 32: 332-336.

[3] S. Ding, X. Zhu, W. Chen, and D. Wei. 2004. Derivation of respiratory signal from single channel ecgs based on source statistics. *Int J of Bioelectromagnetism* 6, 2: 43-49.

[4] A. L. Goldberger, et al. 2000. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation* 101, 23: e215-e220.

[5] T. Magalhães, I. Oliveira, and J. Fernandes. 2015. Message based integration in cyber-physical system: firefighters in the field. In *Proceedings of the MOBIQUITOUS'15, 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 285-286.

[6] B.L. Materna, et al. 1992. Occupational exposures in California wildland firefighting. *Am Ind Hyg Assoc* 53, 1: 69-76.

[7] A.I. Miranda, C. Borrego, and D. Viegas. 1994. Forest fire effects on the air quality. In *Proceedings of the Second International Conference on Air Pollution*, 191-199.

[8] A.I. Miranda, et al. 2005. Smoke measurements during Gestosa 2002 experimental fires. *Int. J. Wildland Fire 14*, 1: 107-116.

[9] A.I. Miranda, et al. 2012. Wildland smoke exposures values and exhaled breath indicator in firefighters. *J. Toxicology and Env. Health 75*, 13-15: 831-843.

[10] G. B. Moody, R. G. Mark, A. Zoccola, and S. Mantero. 1985. Derivation of respiratory signals from multi-lead ECGs. *Computers in Cardiology* 12: 113-116.

[11] J. Mustajbegovic, et al. 2001. Respiratory function in active firefighters. *Am J Ind Med* 40: 55-62.

[12] T.E. Reinhardt, R.D. Ottmar, and C. Castilla. 2001. Smoke impacts from agricultural burning in a rural Brazilian town. *Journal of the Air & Waste Management Association* 51, 3: 443-450.

[13] R. Sebastião, et al. 2016. Inhalation During Fire Experiments: an Approach Derived Through ECG. In Proceedings of the *Ubicomp/ISWC'16 Adjunct* , ACM.

Mixed-Integer Programming Model for the Discovery of Disease Biomarkers Profiles

André M. Santiago[1]
ampacsantiago@student.uc.pt
Miguel Rocha[1]
mrocha@di.uminho.pt
Joel P. Arrais[2]
jpa@dei.uc.pt

[1] Department of Informatics
University of Minho
Braga, Portugal

[2] CISUC
Department of Informatics Engineering
University of Coimbra
Coimbra, Portugal

Abstract

Biomarkers could prove key to understand how we deal with disease diagnostic. For this reason, many methods have emerged over the years that provide means on how to identify potential biomarkers, paving the way for techniques that might help in the diagnostic of disease conditions. We present a mixed-integer linear optimization mathematical model capable of identifying a combination of biomarkers for distinguishing between healthy and diseased samples. This model was tested on two different datasets through sampling analysis, achieving an out of sample accuracy up to 93%.

1 Introduction

The term biological marker, or biomarker for short, was introduced in 1989 as a MeSH (Medical Subject Heading) term, which can be summarized as a biological parameter, measurable and quantifiable in biological samples, which is representative of a specific health or disease state [2]. The discovery of new biomarkers is a very complex process, as a biomarker needs to have high sensitivity and specificity, be reproducibly obtainable through standardized methods, acceptable to the patient in question and easily interpretable by clinical staff [13].

As a general rule, early detection of a disease condition plays a crucial role in successful therapy, which, in most cases, the earlier a disease condition is diagnosed, the more likely it can be successfully cured or well maintained. To this end, the impact and effectiveness that biomarkers may have for diagnostic use has been demonstrated [17]. However, several obstacles still stand in the path for effective recognition of the potential of biomarkers for disease detection [12]: lack of definitive molecular biomarkers for a variety of different diseases, lack of an easy and inexpensive method for sampling and lack of a platform that is portable, accurate and easy-to-use, aiding in early disease detection.

While the field of biomarker discovery has had many interesting developments, there are many different computational technologies available that may yet present themselves as alternative solutions for solving this challenge. Adding to these, an approach is presented in this paper which reduces the biomarker discovery problem to a mathematical model, more specifically a mixed-integer linear optimization model, based on the work of Baliban *et al*. [1] and Puthiyedth *et al*. [15], capable of accurately classifying healthy and disease samples by identifying the optimal combination of biomarkers, while also presenting statistical validation for results obtained from the model.

2 Material and Methods

2.1 Data collection and processing

Two datasets are used for evaluating the mathematical model, one for a condition named esophageal squamous cell carcinoma (ESCC) and another for breast cancer (BC). The first was provided by Su *et al*. [18], with 104 healthy samples and 104 diseased ones, while the second was provided by Maubant *et al*. [14], with a total of 178 samples, of which 11 represent healthy breast tissue. Both datasets went through a filtering process for reducing the number of features to a more manageable size and also selecting the most relevant ones, which represent the most differentially expressed genes. Figure 1 illustrates the distribution of the standard deviation for each dataset's features expression level, after sample-based normalization was performed, which facilitated the task of reducing the number of features and discretizing the expression data, required as the model only considers a gene as present or absent during the optimization procedure. To this end, features whose expression values' standard deviation were below ψ, which is equal to the median of features' standard deviation expression values for each dataset multiplied by γ, which figure 2 illustrates how it was obtained, were removed from each dataset as their variation between healthy and diseased samples was insufficient. Afterwards, to achieve data discretization, for a threshold equal to 3, all expression values equal or above this threshold were rounded to one while the remaining were rounded to zero. This was all achieved through R and Python. The ESCC dataset is available as an ExpressionSet that can be directly loaded through the use of a set of R packages, which include "Category" [7], "GOStats" [8], "affy" [9], "genefilter" [10] and "limma" [16], as well as two annotation packages, "hgu133a.db" [5] and "hgu133b.db" [3], which are all available as part of the Bioconductor project [11]. The BC dataset was available as a SOFT file, requiring only the "GEOquery" [6] R package for loading the data, which is also available as part of the Bioconductor project [11], together with the annotation package "hgu133plus2.db" [4].

2.2 Mixed-integer linear optimization model for prediction of disease profile

A few parameters and variables need to be defined first. Indexes i and j represent a feature and a sample, respectively. H contains all of the healthy samples while D contains all of the diseased ones. y_i and z_j are binary variables, with first being 1 if a feature is selected as a biomarker or 0 when not, while the second is 1 if a sample is classified as diseased while 0 if otherwise. w_i is a continuous variable and represents the weight associated with a feature i. There are also the slack variables, the score error μ_j^+ and the score margin μ_j^- for each sample j. Also, the matrix A represents the data whose expression values were previously discretized, with A_{ij} being equal to 1 if feature i is present in sample j or equal to 0 if not. Furthermore, N represents the total maximum of potential biomarkers considered by the model, I_H and I_D are sets which contain features that have been selected as representative of health or disease condition, w_i^L and w_i^U represent the lower and upper bounds of the weight variables, specific to the feature i, and U is a constant of value equal to 50.

The purpose of the objective function for the mathematical model (1) is to discover the combination of biomarkers for which the sum of their weights is the lowest $\left(\sum_i w_i \cdot A_{ij}\right)$. This means that the selected combination of biomarkers will always try to evaluate all samples as healthy. Adding to this, (1) also favours the minimization of the number of selected features $(0.1 \cdot \sum_i y_i)$, though at a lower priority than determining the best combination of features for the given training set.

$$\min \left(\sum_j^H \left(0.1 \cdot \sum_i y_i - \sum_i w_i \cdot A_{ij} \right) + \sum_j^D \left(0.1 \cdot \sum_i y_i + \sum_i w_i \cdot A_{ij} \right) \right) \quad (1)$$

The model has some constraints to which its variables are subjected to, from (2) to (10).

$$\sum_i w_i \cdot A_{ij} - 1 + \mu_j^+ - \mu_j^- = 0, \quad \forall j \in H \quad (2)$$

$$\sum_i w_i \cdot A_{ij} + 1 - \mu_j^+ + \mu_j^- = 0, \quad \forall j \in D \quad (3)$$

$$\mu_j^+ \geq 0, \quad \mu_j^- \leq 10 \quad (4)$$

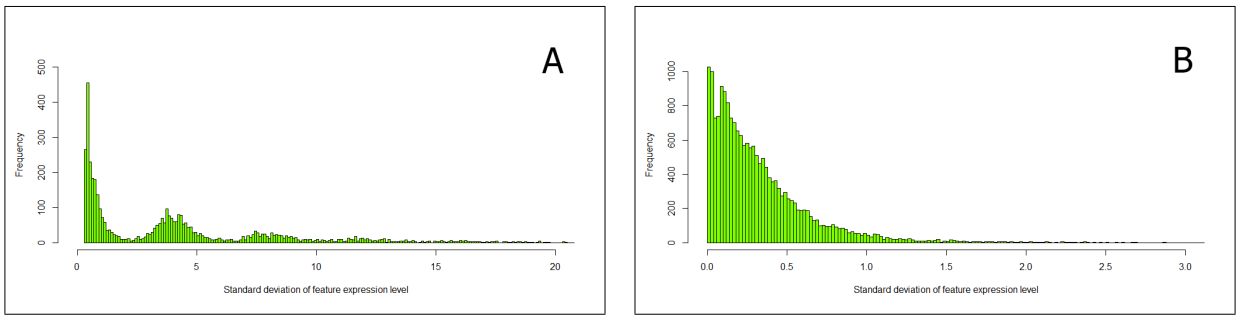


Figure 1: Histograms illustrating the standard deviation distribution of the features' expression level between all the samples for (A) the ESCC dataset and (B) the BC dataset. Features that are more differentially expressed between the samples will have a higher base value.

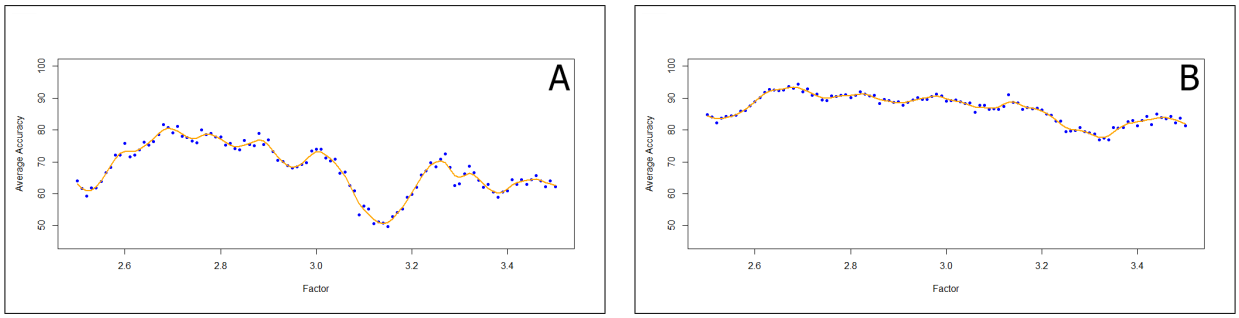


Figure 2: Variation of the accuracy according to the value chosen for γ for the filtering process for (A) the ESCC dataset and (B) the BC dataset. Each accuracy data point on the graphs represents the average of 100 sampling tests with $3/4$ of the samples selected for training and the remaining for testing. The value selected for γ (the one that provided the greatest predictive accuracy) was 2.68 for the former dataset and 2.69 for the latter.

$$\sum_i^{I_H} y_i \geq 1, \quad \sum_i^{I_D} y_i \geq 1, \quad \sum_i y_i \leq N \tag{5}$$

$$y_i \cdot w_i^L \leq w_i \leq y_i \cdot w_i^U, \quad \forall i \tag{6}$$

$$\sum_i^{I_H} w_i \leq \sum_i^{I_H} y_i \tag{7}$$

$$\sum_i^{I_D} w_i \geq -\sum_i^{I_D} y_i \tag{8}$$

$$\sum_i w_i \cdot A_{ij} \geq 1 - z_j \cdot U, \quad \forall j \tag{9}$$

$$\sum_i w_i \cdot A_{ij} \leq (1 - z_j) \cdot U - 1, \quad \forall j \tag{10}$$

All of the equations described above, (1)-(10), constitute a complete mixed-integer linear optimization mathematical model which can be solved to global optimality through the use of CPLEX (ILOG 2013) so as to be able to determine the values of the variables y_i and w_i, which are needed for the scoring function presented below.

2.3 Scoring function.

A function (11) is also proposed for identifying the status of a sample, be it healthy or diseased, by calculating a score for said sample, with $S_j > 0$ implying that a sample is diseased while $S_j < 0$ implies that a sample is healthy. When $S_j = 0$, the condition of the sample is ambiguous.

$$S_j = \lambda \cdot \left(\sum_i w_i \cdot A_{ij} \cdot y_i - \theta \right) \tag{11}$$

With S representing the scores of the samples in the training set T:

$$\theta = mean(S), \quad S \in T \tag{12}$$

$$\lambda = 1000 \cdot std(S), \quad S \in T \tag{13}$$

Table 1: Sampling accuracy for train and test sets, obtained from the ESCC dataset, and how it varies with changes to the size of the train and test sets and to the maximum number of potential biomarkers (N).

# Samples in Training Set	# Samples in Test Set	# Potential Biomarkers (N)	Average Accuracy (Train Set) (%)	Average Accuracy (Test Set) (%)	Average Correct Predictions
52	156	20	76.12	68.37	106.66
		30	77.69	69.22	107.99
		40	79.94	71.94	112.23
104	104	20	80.12	75.37	75.37
		30	83.74	80.33	80.33
		40	84.14	81.18	81.18
156	52	20	84.23	81.90	42.59
		30	86.69	85.17	44.29
		40	86.49	83.50	43.42

Table 2: Sampling accuracy for train and test sets, obtained from the BC dataset, and how it varies with changes to the size of the train and test sets and to the maximum number of potential biomarkers (N).

# Samples in Training Set	# Samples in Test Set	# Potential Biomarkers (N)	Average Accuracy (Train Set) (%)	Average Accuracy (Test Set) (%)	Average Correct Predictions
45	133	20	94.22	86.32	114.81
		30	95.38	85.96	114.33
		40	94.96	85.43	113.62
89	89	20	94.25	90.54	80.58
		30	94.21	90.46	80.51
		40	94.67	90.25	80.32
134	44	20	95.50	93.16	40.99
		30	95.15	91.55	40.28
		40	95.13	92.61	40.75

3 Results and discussion

Biomarker prediction for effective disease diagnostic is a topic which has been expanding in recent years, with many different methods and technologies already available which are able to provide groups of biomarkers that can, usually, accurately distinguish between health and disease cases.

The results obtained through the presented model prove to be quite promising, reaching 93% accuracy on the BC dataset. However, the model is quite sensitive to the data used for training, which figure 2 clearly illustrates. This is more noticeable with the ESCC dataset, which may be due to the fact that the data was obtained from two different, even if similar, array designs.

Moreover, table 1 clearly indicates that a maximum combination of 30 potential biomarkers shows a higher accuracy for the test set than with 20, which may be due to over- or under-fitting, or even 40 biomarkers, in which noise in the expression values may be the culprit. However, for training sets of a smaller size, a maximum combination of 40 potential biomarkers yields the best results, even if only by a very small margin when compared to the results obtained with a maximum of 30 biomarkers. However, table 2 shows that 20 potential biomarkers represents the optimal maximum combination, for the BC dataset, for any training test size, even if the differences between accuracies are quite small, ranging from less than 1% to, at most, 2%.

When compared to other methods present in the literature, the one showed here demonstrates itself as a viable alternative for identifying potential biomarkers. Taking the example of the one developed by Baliban *et al*., from which our methodology is inspired, while the predictive accuracy results presented here are lower than the 99% reported in [1], the type of data used in the latter consists of mass spectrometry data while the one used here consists of gene expression data, which are known to be far more noisier by nature. Adding to this, the dataset used in [1] was also unavailable for comparative analysis.

Another method was developed by Sun *et al*. [19], which identified at the time some potential pairs and even triplets of biomarkers for acute lymphoblastic/myeloid leukemia and colon cancer. While presenting similar predictive accuracy results, the datasets used in [19] considered fewer features and samples in their data, which may have resulted in a larger degree of overfitting. Unfortunately, their datasets were also unavailable to be able to determine how the model developed in this work would perform. Adding to this, the Sun's model is less versatile than ours, considering only a fixed of features in each run while the one showed here is able to consider a varying number of features up to a defined maximum.

Finally, there is also the model developed by Zou *et al*. [20]. Their approach to multi-biomarker panel identification is interesting but suffers from the same issue as Sun's: the number of biomarkers to be identified needs to be fed into the model, while the one presented here takes only into consideration a maximum number, potentially identifying the optimal number and combination of biomarkers. Adding to this, their predictive accuracy results are lower than the ones here demonstrated, reaching only up to 89% accuracy.

References

[1] Richard C Baliban, Dimitra Sakellari, Zukui Li, Yannis A Guzman, Benjamin A Garcia, and Christodoulos A Floudas. Discovery of biomarker combinations that predict periodontal health or disease with high accuracy from GCF samples based on high-throughput proteomic analysis and mixed-integer linear optimization. *Journal of clinical periodontology*, 40(2):131–9, mar 2013. ISSN 1600-051X. doi: 10.1111/jcpe.12037.

[2] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology and therapeutics*, 69(3):89–95, mar 2001. ISSN 0009-9236. doi: 10.1067/mcp.2001.113989.

[3] Marc Carlson. *hgu133b.db: Affymetrix Human Genome U133 Set annotation data (chip hgu133b)*. R package version 3.1.3., .

[4] Marc Carlson. *hgu133plus2.db: Affymetrix Human Genome U133 Plus 2.0 Array annotation data (chip hgu133plus2)*. R package version 3.1.3., .

[5] Marc Carlson. *hgu133a.db: Affymetrix Human Genome U133 Set annotation data (chip hgu133a)*. R package version 3.1.3, .

[6] Sean Davis and Paul S Meltzer. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics (Oxford, England)*, 23(14):1846–7, jul 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btm254.

[7] R. G. Falcon and D. Sarkar. *Category: Category Analysis.* R package version 2.34.2.

[8] S Falcon and R Gentleman. Using GOstats to test gene lists for GO term association. *Bioinformatics (Oxford, England)*, 23(2):257–8, jan 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btl567.

[9] Laurent Gautier, Leslie Cope, Benjamin M Bolstad, and Rafael A Irizarry. affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics (Oxford, England)*, 20(3):307–15, feb 2004. ISSN 1367-4803. doi: 10.1093/bioinformatics/btg405.

[10] R Gentleman, V Carey, W Huber, and F Hahne. *genefilter: methods for filtering genes from microarray experiments*. R package version 1.50.0.

[11] Robert C Gentleman, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, Kurt Hornik, Torsten Hothorn, Wolfgang Huber, Stefano Iacus, Rafael Irizarry, Friedrich Leisch, Cheng Li, Martin Maechler, Anthony J Rossini, Gunther Sawitzki, Colin Smith, Gordon Smyth, Luke Tierney, Jean Y H Yang, and Jianhua Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80, jan 2004. ISSN 1465-6914. doi: 10.1186/gb-2004-5-10-r80.

[12] Yu-Hsiang Lee and David T Wong. Saliva: an emerging biofluid for early detection of diseases. *American journal of dentistry*, 22(4): 241–8, aug 2009. ISSN 0894-8275.

[13] Teri Manolio. Novel risk markers and clinical practice. *The New England journal of medicine*, 349(17):1587–9, oct 2003. ISSN 1533-4406. doi: 10.1056/NEJMp038136.

[14] Sylvie Maubant, Bruno Tesson, Virginie Maire, Mengliang Ye, Guillem Rigaill, David Gentien, Francisco Cruzalegui, Gordon C Tucker, Sergio Roman-Roman, and Thierry Dubois. Transcriptome analysis of Wnt3a-treated triple-negative breast cancer cells. *PloS one*, 10(4):e0122333, jan 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0122333.

[15] Nisha Puthiyedth, Carlos Riveros, Regina Berretta, and Pablo Moscato. A New Combinatorial Optimization Approach for Integrated Feature Selection Using Different Datasets: A Prostate Cancer Transcriptomic Study. *PloS one*, 10(6):e0127702, 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0127702.

[16] M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):47, jan 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv007.

[17] Anne-Sofie Schrohl, Sidse Würtz, Elise Kohn, Rosamonde E Banks, Hans Jørgen Nielsen, Fred C G J Sweep, and Nils Brünner. Banking of biological fluids for studies of disease-associated protein biomarkers. *Molecular & cellular proteomics : MCP*, 7 (10):2061–6, oct 2008. ISSN 1535-9484. doi: 10.1074/mcp.R800010-MCP200.

[18] Hua Su, Nan Hu, Howard H Yang, Chaoyu Wang, Mikiko Takikita, Quan-Hong Wang, Carol Giffen, Robert Clifford, Stephen M Hewitt, Jian-Zhong Shou, Alisa M Goldstein, Maxwell P Lee, and Philip R Taylor. Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 17(9):2955–66, may 2011. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-10-2724.

[19] Minghe Sun and Momiao Xiong. A mathematical programming approach for gene selection and tissue classification. *Bioinformatics (Oxford, England)*, 19(10):1243–51, jul 2003.

[20] Meng Zou, Peng-Jun Zhang, Xin-Yu Wen, Luonan Chen, Ya-Ping Tian, and Yong Wang. A novel mixed integer programming for multi-biomarker panel identification by distinguishing malignant from benign colorectal tumors. *Methods (San Diego, Calif.)*, 83:3–17, jul 2015. ISSN 1095-9130. doi: 10.1016/j.ymeth.2015.05.011.

A practical study about the Google Vision API

Daniel Pedro Ferreira Lopes
lopesdaniel@ua.pt
António J. R. Neves
http://sweet.ua.pt/an/
Universidade de Aveiro, Portugal

Abstract

Google Vision API is an application programming interface provided by Google that enables developers to understand the content of an image. Due to its powerful machine learning models and a huge database of images, not only it quickly classifies images into thousands of categories, but also detects individual objects and faces within images.

In this paper we present a practical study regarding the use of this library in order to studied some of the Google Vision features to check their usability in practical applications.

1 Introduction

In everyday life there are many problems that can be solved through computer vision solutions and some of the projects under development at the Institute of Electronics and Informatics Engineering of Aveiro (IEETA) are inserted on this category. Problems like face recognition, emotional states and image classification can be solved with some of the market solutions available. Google Vision API [1] and Microsoft Cognitive Services [2] are some low-cost answers that are currently on the market. The current state of art tells that object recognition can be made through Selective Search [4] or Convolutional Networks [3]. Also, face recognition is made either with Fisherfaces or EigenFaces [5]. Google takes these algorithms and applies its machine learning processes to improve them.

In this work we evaluated the Google Vision API in order to realize what solutions may be used to current or future projects under development at IEETA by trying to perceive its potential and fails.

2 Features

These are the currently detections features available on the API:

- Label Detection
- Explicit Content Detection
- Logo Detection
- Landmark Detection
- Optical Character Recognition
- Face Detection

In this paper we studied Label, Landmark and Face detection.

2.1 Label Detection

Label Detection feature detects broad sets of categories within an image, ranging from modes of transportation to animals. The response give us various results along with a confidence level (between 0 and 1).

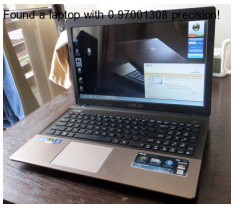


Figure 1: Picture of a Laptop

Figure 2: Picture of a Pen

Figure 3: Picture of a Wallet

Figure 4: Picture of a Calculator

For the request made with Figure 1, Google API responded "Found a laptop with 0.97 score.". As for Figure 2 it says "Found a pen with 0.82 score.". Both of the responses are quite accurate and with a high score.

For Figure figure 3 Google says: "Wallet with a score of 0.75". However, Google got wrong when it says that Figure 4 is a "Phone with a score of 0.77".

Figure 5: Picture of a Piano

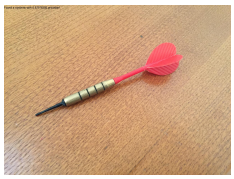


Figure 6: Picture of a Dart

For Figure figure 5 Google says: "Piano with a score of 0.96" which is a very accurate result. However, Google gets totally wrong when it says that figure 6 is a "Eyebrow with a score of 0.87".

We tested more images and almost all the results were accurate.

There were requested 9 photos taken with a camera and 12 Google Images of some everyday items in diverse environments. The results are presented on Table 1.

	Google Pictures	Camera Pictures
Not Accurate	3	3
Sort of Accurate	2	2
Really Accurate	7	4
Total	12	9

Table 1: Classification of Google's responses for label detection

We can notice that, in general, the image labelling works well. It is noteworthy that, comparing the requests made with Google Images are much more accurate responses than the images taken with the camera. This phenomenon is most likely occurred because the images that are found on Google are the ones that are used for Google's machine learning and consequently the images' data base that are improving the API's algorithm everyday.

2.2 Landmark Detection

The Landmark Detection feature detects popular natural and man-made structures within an image.

Figure 7: Photo of New York City

Figure 8: Regions where Google bases to get the response

For Figure 7 Google API's response was: "Streets from New York". In the response it also gives the bounding boxes where it bases to get that response as it shows in Figure 8.

Figure 9: Photo of Forbidden City, Beijing

Figure 10: Region where Google bases to get the response

Goggle's response for Figure 9 was: "Forbidden City, Hall of Supreme Harmony". The result was accurate.

There were requested 4 photos from Google Images and 4 images taken from a camera. The results are shown on Table 2.

	Google Pictures	Camera Pictures
No response	0	2
Wrong response	0	0
Right response	4	2
Total	4	4

Table 2: Classification of Google's responses for Landmark Detection

Landmark detection algorithm is very precise, specially with images already indexed by Google. From the images requested, either it gives a correct answer or does not give nothing at all. There is no incorrect answers. Unlike label detection, landmark detection does not give the score of each answer. It is also observable that the monuments or landscapes that are more famous are easily detected while the less ones (like an Aveiro University photo) are not. As Google image database is getting bigger, the landmark detection is also improving.

2.3 Face Detection

The Face Detection feature detects multiple faces within an image, along with the associated key facial attributes like emotional state or wearing headwear. In this feature, **Facial Recognition is not supported.**

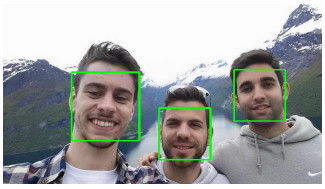


Figure 11: Faces recognized. Photo taken by a cellphone camera

We tested 20 photos with one or more faces on it. These results are shown on Table 3.

	One face	Two faces	Three or more faces
None detected	1	0	0
Some detected	0	0	1
All detected	12	3	3
Total	13	3	4

Table 3: Face detection performance

From Table 3, it is possible to notice that face detection had 90% of accuracy, with this tested dataset. However, it is not possible to classify the accuracy rate of the emotional state of each person. That is because it is impossible to take the Google's responses and put them on a graph or a table.

Face detection + Emotional State:

It was also explored the potential of Google to analyse the emotional states.

As it is noticeable both of the faces were detected. As for emotional states Google tells that on Figure 12 that it is very likely that the person shows joy and very unlikely that shows anger or surprise. In the Figure 13

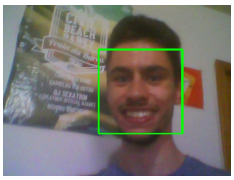


Figure 12: Person smiling

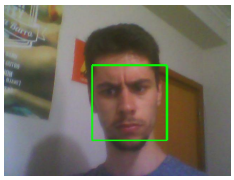


Figure 13: Person angered

Google tells that it is unlikely that the person is angered and it is very unlikely that he is joyed or surprised.

Conclusions

Face Detection algorithm used by Google Vision API is very powerful and almost flawless. As it was tested in several photos with a single or multiple faces, there were detected almost all the times. When it comes for emotional states interpretation, it is possible to say that Google has some work on that aspect. When a person is smiling it automatically detects joy. However, when someone is angered or surprised, Google still struggles to interpret that states.

3 Final Conclusions and Remarks

After this study it is conclusive that Google Vision API has a great potential and it is getting better as the time passes. Its powerful Machine Learning and huge image database helps to be one of the best current market solutions.

Landmark Detection has a promising future in labelling photos taken during a trip and organized them in albums.

Regarding to Face Detection, it will be helpful in future research on people's emotions when exposed to a certain environment or a certain product.

Finally when it comes to Label Detection, it helps to analyse a picture and say what is on it. This would be helpful for future robotic and computer vision systems.

There is still an important disadvantage regarding this API. It can not be used in real-time systems since the communication with the Google's servers introduces a considerable delay.

In conclusion, Google Vision API can be seen as an interesting alternative solution to some of the current IEETA projects.

4 References

[1] Google Vision API, *https://cloud.google.com/vision/*

[2] Microsoft Cognitive Services, *https://www.microsoft.com/cognitive-services/*

[3] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, Yann LeCun, *Integrated Recognition, Localization and Detection using Convolutional Networks* 2014.

[4] K E A van de Sande, J.R.R. Uijlings, T Gevers, A.W.M. Smeulders, *Segmentation as Selective Search for Object Recognition* 2011.

[5] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection* 1997.

Estimation of choroidal thickness in OCT images

Simão P. Faria [3]
Susana Penas [4]
Luís Mendonça [5]
Jorge A. Silva [2,3]
Ana Maria Mendonça[1,3]

1- Dep. of Electrical and Computer Eng., Faculty of Engineering, University of Porto
2- Dep. of Informatics Engineering, Faculty of Engineering, University of Porto
3- INESC TEC, Porto
4- Dep. of Ophthalmology, São João Hospital Center
5- Dep. of Ophthalmology, Hospital of Braga

Abstract

The choroid is the middle layer of the eye globe located between the retina and the sclera. It is proven that choroidal thickness is a sign of multiple eye diseases. Optical Coherence Tomography (OCT) is an imaging technique that allows the visualization of tomographic images of near surface tissues like those in the eye globe. The automatic calculation of the choroidal thickness reduces the subjectivity of manual image analysis as well as the time of large scale measurements.

In this paper, a method is presented for the automatic estimation of the choroidal thickness in OCT images. The pre-processing of the images is focused on noise reduction, shadow removal and contrast adjustment. The inner and outer boundaries of the choroid are delineated sequentially, resorting to a minimum path algorithm. The choroidal thickness is given by the distance between the two boundaries.

The method was evaluated by calculating the error as the absolute distance from the automatically estimated outer boundaries to the boundaries delineated by two ophthalmologists, in 14 images. The differences between the two sets of manual boundaries are usually larger than the error of the automatic segmentation.

Usually OCT scans are performed in two perpendicular sets, allowing a comparison of the segmentation of one set with the other, after alignment and interpolation. The differences in choroidal thickness measured from each set have a median of approximately 2.2% of the average thickness.

1 Introduction

The choroid is the middle layer of the eye globe located between the retina and the sclera (Figure 1), bordered internally by the Bruch's Membrane (BM) and externally by the Choroidal-Scleral Interface (CSI). Its purpose is to provide metabolic needs to the retina and to regulate ocular pressure and temperature[3]. The vascular nature of the choroid makes the variations of its thickness an indication for the ocular health. Pathologies as retinitis pigmentosa, serous chorioretinopathy, diabetic retinopathy and others that cause inflammation in the tissue may induce the thickening of the choroid, while the narrowing can be associated with myopia, dehydration or age [2].

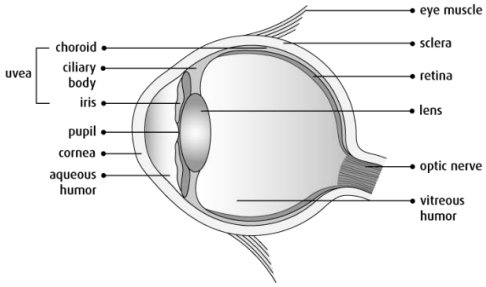


Figure 1: Anatomical model of the eye (image source:[4]).

The OCT is a non-invasive, non-painful, fast imaging technique that can get information unavailable through other imaging methods. This tomographic technique enables the in vivo visualization of subsuperficial tissues with high resolution. Developing software tools to help physicians get the most of this recent technique will certainly contribute for the health of multiple patients.

Recent technological advances in the optical OCT, allowed a better visualization of deeper structures such as the CSI, so the automatic segmentation of the choroid becomes more viable [2].

2 Methods

2.1 Automatic Segmentation

In order to estimate the thickness of the choroid, its two boundaries, the BM and the CSI, must be detected (Figure 2). An additional curve, named Interdigitation Zone (IZ), corresponding to a hyperreflective layer inside the retina (Figure 2) must also detected, in a preliminary step. This curve is used as a reference for the delineation of the BM. In the

The IZ, BM and CSI curves are determined using a minimum cost path algorithm (MCPA) [3], based on a cost function that depends on the characteristics of the curve to delineate. This cost function is expressed by a cost matrix that has the size of the image it is generated from.

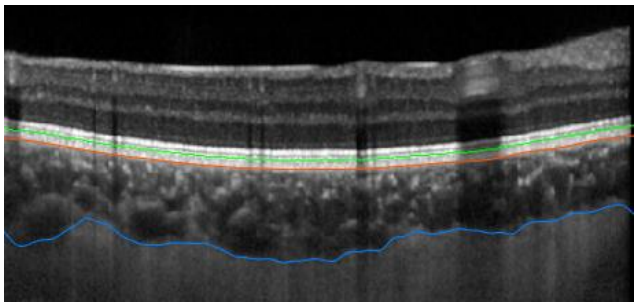


Figure 2: OCT image with the detected curves: Interdigitation Zone (IZ), in green; Bruch's Membrane (BM), in orange; and Choroidal-Scleral Interface (CSI), in blue. The choroid is delimited by BM and CSI.

In a first step, the contrast of the image is adjusted by raising the value of each pixel to the fourth power [4]. This allows an enhancement of the brighter areas of the retina (Figure 3.a) that is fundamental for the detection of the IZ.

The cost matrix for the delineation of the IZ, using the MCPA, is obtained by calculating the negative image of a smoothed version of the contrast adjusted image [1]. An average filter, with 60 by 55 μm, is used for the smoothing, that agrees with the normal thickness of the brightest region in the healthy eye (approximately 60 μm).

The estimation of the IZ will aid the location of the BM. The cost matrix for the delineation of the BM incorporates two penalty terms: one for paths that are far away from the IZ and another one for paths above the IZ. On the other way, a term is included in the cost matrix that favours the curves that are in the transition from the brighter layers of the retina to the darker choroidal tissue; this term is based on the value of a Sobel derivative [3].

After delineating the BM, another contrast compensation algorithm is applied to the image. The objective of this processing is to compensate for the loss of the signal energy in the ocular tissue and to deal with the shadows cast by retinal vessels [1][3]. The resulting image (Figure 3.b) has a better contrast in the CSI region; however the speckle noise is enhanced.

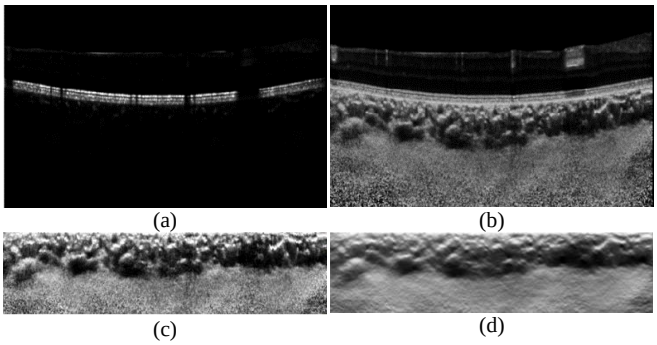
The following step is the flattening and cropping of the image [3] (Figure 3.c). The flattening is done by shifting every column of the image so that the BM becomes a horizontal straight line. Cropping is done by cutting the image 515 μm below the BM; this height guarantees that the choroid is included in the cropped image. This step is made to limit the path search to a region of interest.

In order to reduce the noise and attenuate the shadows, before the delineation of the CSI, a stationary wavelet transform is applied. A five level wavelet decomposition is done, using a Haar wavelet [5], and a hard threshold is applied to the horizontal coefficients of the decomposition. By setting this threshold to zero, the resulting image has less horizontal details, which helps to attenuate shadows (Figure 3.d) and keeps the intensity of the vertical transitions between the choroid and the sclera.

Finally, to delineate the CSI, the cost matrix used in the MCPA is based on the sum of two components that favour paths in the transition between the darker parts of the choroid and the lighter sclera. The first one is computed by applying a morphological opening (to remove the white parts of the choroid) with an 80 by 60 μm kernel, followed by Gaussian smoothing and vertical derivative computation, with a vertical length of 512 μm. The second one uses the same kind of smoothing and derivative computation, but with a smaller kernel, for a fine localization, and is only defined in the dark-to-light transitions identified by the first component [1]. The use of two kernels allows a good spatial

discrimination without a great impact of the choroidal vessels and noise that can have the same type of transitions as the CSI and generate errors.

The thickness of the choroid is calculated as the mean vertical distance between the BM and the CSI.



(a) (b)

(c) (d)

Figure 3: OCT image after: (a) contrast adjustment; (b) compensation algorithm; (c) flattening and cropping; (d) wavelet based filtering.

2.2 Merging sets of B-scans

Usually two sets of OCT scans are acquired for the same eye, in perpendicular directions. Each set consists of 49 parallel B-scans, covering a rectangular area. Figure 4 shows two infrared (IR) images of the fundus of an eye, where the covered area is signalled; the arrowed line indicates the position and the direction of one of the acquired B-scans

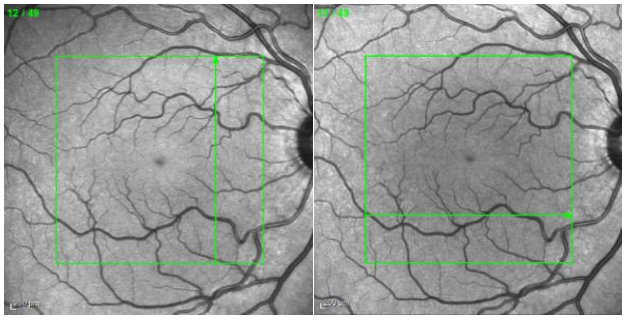
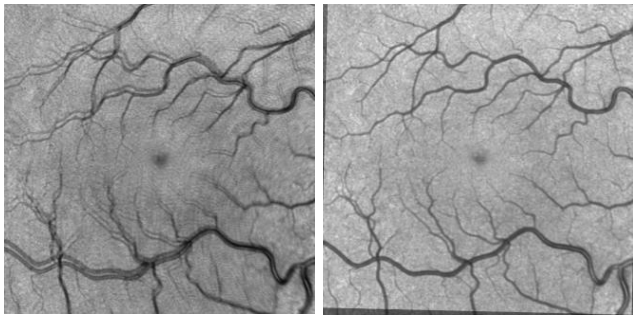


Figure 4: Examples of the IR images with the location of the B-scans.

It is possible to improve the thickness measures by using the information obtained from both sets. However, originally they are not anatomically aligned. An image registration algorithm is used in order to align the IR images. The algorithm based in the maximization of mutual information by translating and rotating the images (Figure 5). The resulting transformation matrix is used for the alignment of the B-scan data.



(a) (b)

Figure 5: Example of the overlapping IR images that represent the area scanned by the OCT. (a) original IR images of two OCT videos of the same eye are merged in one representation; (b) the IR images are aligned using the image registration algorithm.

For the comparison between measures obtained from the two sets, an interpolated 3D surface representing the CSI is obtained for each set. This surface is sampled at each one of the positions of the orthogonal B-scans, obtaining interpolated CSI boundaries that can be compared with the boundaries detected in those orthogonal scans (Figure 6).

The comparison of the surfaces obtained from the two orthogonal scans allowed an additional evaluation of the algorithm's precision.

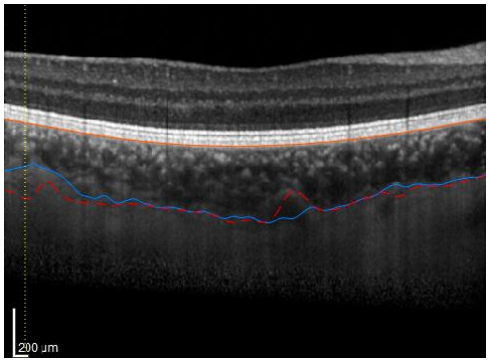
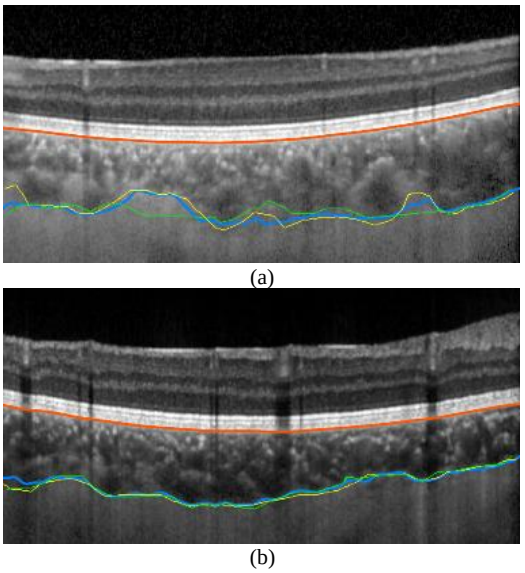


Figure 6: Example of an OCT B-scan with plots of the choroidal boundaries. In orange, the BM; in blue the CSI; the dotted red line represents the interception of the CSI membrane calculated using the series of B-scans in the perpendicular direction; the yellow dotted line locates the maximum difference between the blue and the dotted red lines.

3 Results and Discussion

To evaluate the performance of the automatic delineation of the CSI, the results of the algorithm were compared with manual markings delineated by two ophthalmologists, in 14 OCT images (7 horizontal and 7 vertical) of a single eye. This also allowed an analysis of the interobserver variability.

As observable in Figure 7, there are situations where the similarity between the three markings is clearly visible (Figure 7.a) and others where even the manual markings have discrepancies in certain zones (Figure 7.b) caused by the inclusion of vessels from the proximal sclera.



(a)

(b)

Figure 7: OCT images with the automatically detected BM (orange) and CSI (blue), and the two manual markings (yellow and green).

Results show that the errors for the automatic segmentation are comparable to the differences between the manual paths. The average errors are 12.3±12.6 µm and 12.6±13.7 µm, for each one of the ophthalmologists' markings. These errors are lower than the average difference between the two manual paths: 14.4±16.5 µm.

The 3D interpolated CSI boundaries, were used for the calculation of the differences between the choroidal thicknesses estimated from each set of B-scans. The histogram of the differences (Figure 8) shows that they are predominantly small. The mean absolute difference is approximately 11.49 μm and the median is 6.45 μm, meaning that the majority of differences are smaller than 2.2% of the average choroidal thickness of the analyzed eye (288.54 μm).

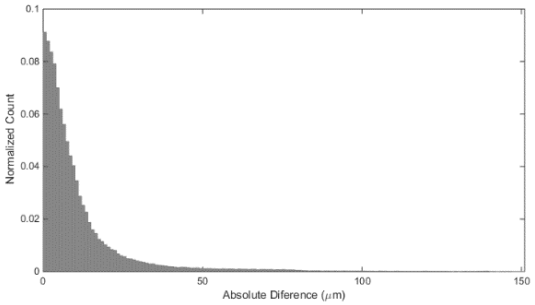


Figure 8: Histogram of the values of the thickness differences between the two sets of B-scans. The normalized count is the number of points divided by the total number of common points.

The differences in thickness measured from the two orthogonal sets can be caused by the way some vessels are sectioned in the tomographic image: in one direction they may seem as part of the choroid while in the other they may seem as scleral vessels. This kind of errors can also cause an incorrect manual delineation, because in common OCT analysis, the ophthalmologist does not make this kind of comparison between different sets of B-scans.

In the developed software application, the CSI interface obtained from each set is shown to the user, who has the possibility of manually correcting the automatic delineations.

4 Conclusion

A method for automatically estimating the choroidal thickness in OCT images was described. The choroidal thickness was calculated by detecting the internal and external boundaries (the BM and the CSI).

The dataset of OCT images allowed an interpolation of the delineated boundaries to obtain 3D surfaces. It was possible to align the B-scans and compare the segmentations in two orthogonal directions.

The final results of the automatic choroidal segmentation were very adequate. The errors in the position of the automatic boundaries are lower than the differences between the two manual markings made by physicians.

The developed software was tested on 98 B-scans. In future development of this work, the robustness of this algorithm should be evaluated using a larger set with ground truth BM and CSI boundaries, from different patients.

Acknowledgements

This work is supported by Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] S. P. Faria, “Estimation of choroidal thickness in OCT images,” Faculty of Engineering, University of Porto, 2016.

[2] A. González-López, B. Remeseiro, M. Ortega, and M. G. Penedo, “Choroid characterization in EDI OCT retinal images based on texture analysis,” *ICAART 2015 - 7th Int. Conf. Agents Artif. Intell. Proc.*, vol. 2, pp. 269–276, 2015.

[3] D. Alonso-Caneiro, S. A. Read, and M. J. Collins, “Automatic segmentation of choroidal thickness in optical coherence tomography.,” *Biomed. Opt. Express*, vol. 4, no. 12, pp. 2795–812, Dec. 2013.

[4] “http://www.cancer.ca/en/cancer-information/cancer-type/eye/anatomy-and-physiology/?region=on.”, accessed on Sep 2016.

[5] H. Danesh, R. Kafieh, H. Rabbani, and F. Hajizadeh, “Segmentation of choroidal boundary in enhanced depth imaging OCTs using a multiresolution texture based modeling in graph cuts,” *Comput. Math. Methods Med.*, vol. 2014, Article ID 479268, 9 pages, 2014.

Segmentation of the Left Ventricle in Cardiac MRI using a Robust Active Shape Model Approach

Carlos Santiago
carlos.santiago@ist.utl.pt
Jacinto C. Nascimento
jan@isr.ist.utl.pt
Jorge S. Marques
jsm@isr.ist.utl.pt

Instituto de Sistemas e Robótica,
Instituto Superior Técnico,
Lisboa, Portugal

Abstract

The segmentation of the left ventricle in MRI is a required task to evaluate and diagnose cardiac function. The main challenges for the development of an automatic segmentation tool are: i) the presence of misleading anatomical structures, ii) edge fuzziness near the apical and basal slices, and iii) misalignment between consecutive slices. A common approach is to use shape information to guide the segmentation, e.g., using an Active Shape Model (ASM). However, the presence of outliers hampers the accuracy of this approach. This paper proposes an EM framework that takes outliers into account and is able to provide robust segmentation estimates. The proposed method was evaluated on a public dataset with 33 MR sequences and the results show it provides significant improvements over the standard ASM method, and also outperforms another state of the art approach.

1 Introduction

Cardiac MRI is the standard image modality for the assessment and diagnosis of some cardiomyopathies [4]. After acquiring a sequence of MR volumes, covering a entire cardiac cycle, cardiologists have to manually delineate the inner boundary of the left ventricle (LV), called endocardium. Only then are they able to compute specific features of cardiac function, such as ventricle volumes and ejection fraction.

To relieve cardiologists of this morose task, several (semi)automatic segmentation algorithms have been proposed over the years [5]. However, automatically identifying the endocardium is a complex task, due to: 1) wall irregularities, caused by the presence of papillary muscles and trabeculations; 2) edge fuzziness near the apical and basal slices, due to partial volume effects, and 3) misalignment between consecutive volume slices that may appear due to different breath-holding levels during acquisition.

A popular approach is to use shape priors to constrain the final segmentation [3]. Among this type of approaches, one of the most popular is the Active Shape Model (ASM) [2]. This method uses an explicit representation of the contour that is able to deform according to specific modes of variation observed in a training set of annotated data.

Although ASM based methods have achieved state of the art results, their performance is often hampered by the presence of other misguiding boundaries, typically denoted as outliers. This paper proposes a robust ASM that is able to achieve accurate results even in the presence of outliers [6]. The proposed method is based on an Expectation-Maximization (EM) approach that assumes each edge segment detected in the MR volume may either belong to the endocardium or to outliers. Under this assumption, each edge segment is assigned a specific weight during the segmentation process depending on the probability that it belongs to the LV boundary, as will be explained in the following section.

2 Proposed Methodology

2.1 Shape Model

In this work, a 3D shape model is used to define the segmentation of each MR volume. This shape model is learned using the approach described in [7], which provides a framework to overcome the challenge of learning a 3D shape model from annotated volumes with a variable number of slices. Formally, it allows a specific slice model, $\boldsymbol{x}(s) = \left[\boldsymbol{x}^1(s),\ldots,\boldsymbol{x}^N(s)\right] \in \mathbb{R}^{2N}$, to be obtained from the training set, where $s \in [0,1]$ denotes the position of that slice in the MR volume, and $\boldsymbol{x}^j(s) \in \mathbb{R}^2$ defines the position of the j-th model point within the slice. The position of the model points is defined by the set of parameters, $\boldsymbol{\theta} = \{\boldsymbol{a},\boldsymbol{t}\}$ and $\boldsymbol{b}(s)$, which correspond to the parameters of the similarity (pose) transformation and the deformation coefficients, respectively, through

$$\boldsymbol{x}(s) = \boldsymbol{T}_{\boldsymbol{\theta}}\left(\bar{\boldsymbol{x}}(s) + \boldsymbol{D}(s)\boldsymbol{b}(s)\right), \tag{1}$$

where $\bar{\boldsymbol{x}}(s) \in \mathbb{R}^{2N}$ is the average model in slice position s, $\boldsymbol{D}(s) \in \mathbb{R}^{2N\times L}$ is the matrix of deformation modes, and $\boldsymbol{T}_{\boldsymbol{\theta}}$ is the linear transformation that defines the pose of the LV.

Unlike typical ASMs, the shape model used in this work includes two additional deformation modes, which grant each slice model the ability to move within the slice plane, thus allowing them to fit misaligned slices.

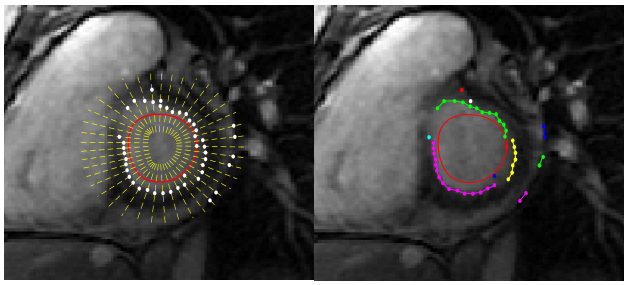


Figure 1: Extraction of candidate edge segments from a volume slice: (left) detection of edge points (white) along lines (yellow) orthogonal to the model (red); (middle) edge segments obtained by linking edge points.

2.2 EM Framework

The goal of the algorithm is to segment all the slices of a particular MR test volume. Suppose this volume has S slices, and that the position of the m-th slice is given by s_m. In order to fit the shape model to the boundary of the LV in all the slices, the model parameters, $\boldsymbol{\theta} = \{\boldsymbol{a},\boldsymbol{t}\}$ and $\boldsymbol{b}(s_1),\ldots,\boldsymbol{b}(s_S)$, have to be chosen accordingly. Candidate points are extracted from each slice by searching along lines orthogonal to the model, as depicted in Fig. 1 (left). Then, edge segments are obtained by grouping these candidates, as shown in Fig. 1 (right).

Each of the detected edge segments, denoted by $\boldsymbol{Y}^i(s_m) \in \mathbb{R}^{2M^i}$, may belong to the endocardium or to outliers. Since this information is not known *a priori*, a binary hidden variable, $k^i(s_m)$, is used to allow both possibilities: $k^i(s_m) = 1$ for valid segments and $k^i(s_m) = 0$ for outliers. These two possibilities are assumed to occur with probabilities p_1 and p_0, and, for each case, a different observation model is used, as follows

$$p\left(\boldsymbol{Y}^i(s_m)\middle|k^i(s_m)=1,\boldsymbol{\theta}\right) = \prod_{j=1}^{M^i}\mathcal{N}\left(\boldsymbol{y}^{ij}(s_m);\boldsymbol{x}^{ij}(s_m),\boldsymbol{\Sigma}^{ij}(s_m)\right), \tag{2}$$

$$p\left(\boldsymbol{Y}^i(s_m)\middle|k^i(s_m)=0,\boldsymbol{\theta}\right) = \prod_{j=1}^{M^i}\mathcal{U}(V_{\boldsymbol{x}^{ij}(s_m)}), \tag{3}$$

where $\boldsymbol{y}^{ij}(s_m) \in \mathbb{R}^2$ is the j-th candidate point in $\boldsymbol{Y}^i(s_m)$ and $\boldsymbol{x}^{ij}(s_m)$ is the corresponding model point. $\mathcal{N}(\cdot)$ and $\mathcal{U}(\cdot)$ define a Gaussian and a uniform distribution, respectively. $\boldsymbol{\Sigma}^{ij}(s_m)$ is the variance associated with the corresponding model point, $\boldsymbol{x}^{ij}(s_m)$, and $V_{\boldsymbol{x}^{ij}(s_m)}$ defines a validation gate in the vicinity of $\boldsymbol{x}^{ij}(s_m)$.

Let $\boldsymbol{\Theta} = \{\boldsymbol{a},\boldsymbol{t},\boldsymbol{b}(s_1),\ldots,\boldsymbol{b}(s_S),p_1,p_0\}$ define the complete set of parameters, and let $\mathcal{Y}$ and $\mathcal{K}$ be the set of all the detected segments in all

This work was supported by FCT [SFRH/BD/87347/2012] and [UID/EEA/50009/2013].

the slices and their corresponding labels. The EM framework allows $\boldsymbol{\Theta}$ to be iteratively optimized by maximizing the expectation of the complete posterior probability,

$$\widehat{\boldsymbol{\Theta}}_{(t+1)} = \arg\max_{\boldsymbol{\Theta}} Q\left(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}_{(t)}\right) = \mathbb{E}_{\mathcal{K}}\left[\mathcal{P}(\mathcal{Y}, \mathcal{K}, \boldsymbol{\Theta}) \middle| \mathcal{Y}, \widehat{\boldsymbol{\Theta}}_{(t)}\right], \quad (4)$$

in a two step procedure. In the first step, *E-step*, the probability of each edge segment being valid (or outlier) is updated based on the current model estimate

$$w_1^i(s_m) = p\left(k^i(s_m)=1 \middle| \boldsymbol{Y}^i(s_m), \widehat{\boldsymbol{\Theta}}_{(t)}\right) \quad (5)$$

$$w_0^i(s_m) = p\left(k^i(s_m)=0 \middle| \boldsymbol{y}^{ij}(s_m), \widehat{\boldsymbol{\Theta}}_{(t)}\right), \quad (6)$$

such that $w_1^i(s_m) + w_0^i(s_m) = 1$. In the second step, *M-step*, the model parameters are updated by maximizing (4). This leads to a weighted least squares regression that minimizes the distance between the model points and the corresponding edge segments, where each segment is weighted by (5).

Since outliers typically receive lower weights, their influence in the estimation of the model parameters is reduced, leading to more robust results. The algorithm iterates between the two steps until the parameters converge.

3 Results

The proposed algorithm was evaluated on a public dataset [1], which contains 33 MR sequences, each with 20 volumes. The results were obtained using a leave-one-sequence-out scheme: for each test sequence, the shape model was learned using the remaining 32 sequences.

The segmentations were evaluated by comparison with the ground-truth using two metrics: 1) the Dice coefficient, which measures the agreement between two segmented regions, and 2) the average distance (AV) between the model points and the ground-truth. Statistical results for another state of the art approach [6] are also provided for comparison.

Fig. 2 shows some examples of the segmentations obtained using the proposed method. It is possible to see that the proposed segmentation is similar to the ground-truth in most cases. Fig. 3 shows other examples with a color-coded evaluation of each slice segmentation. Two conclusions can be drawn. First, volumes in the end-dyastolic frame are easier to segment, which is expected since the LV is dilated and its borders are more noticeable. Second, the segmentation of the apical slice is typically poorer than the remaining slices, due to the partial volume effect mentioned in Section 1.

Table 1 summarizes the statistical results obtained using the proposed approach and using the RANSAC algorithm proposed in [6]. The results show that the proposed method outperforms the RANSAC approach, even though the latter also brings significant improvements over the standard ASM proposed in [2].

Table 1: Statistical performance of the proposed algorithm (mean and standard deviation) and comparison with other approaches.

	Dice (%)	AV (mm)
ASM [2]	73.1 (13.1)	4.7 (3.0)
RANSAC [6]	83.0 (7.5)	2.7 (1.0)
Proposed	**85.8 (6.7)**	**2.2 (0.6)**

4 Conclusion

This paper proposes a robust Active Shape Model approach that is able to deal with the difficulties associated with cardiac MRI analysis: the presence of other anatomical structures that misguide the model (outliers) and the existence of misaligned slices. The proposed approach is based on an EM framework that takes into account the presence of outliers. By assigning weights to candidate edge segments extracted from the image, the algorithm is able to reduce the influence of outliers in the estimation of the model parameters, thus leading to robust segmentations. Significant improvements over the standard ASM and another state of the art approach are achieved and show that the proposed method is able to provide good LV segmentations.

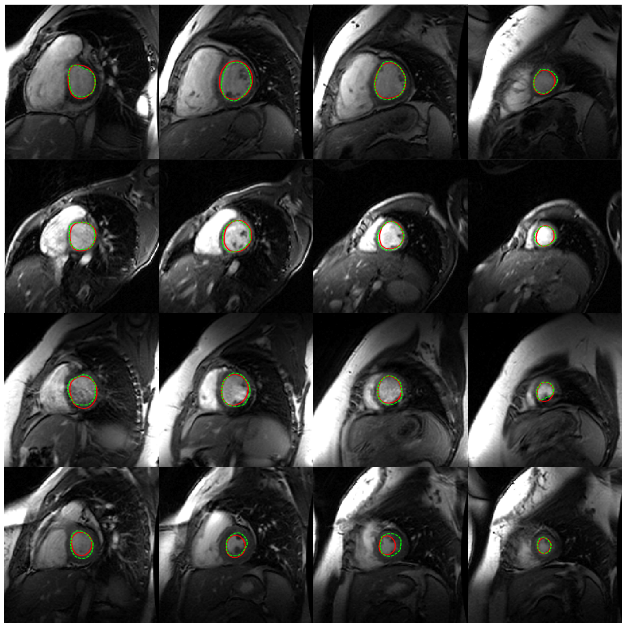


Figure 2: LV segmentation. Each row shows four slices of a different volume depicting: the segmentation obtained using the proposed method (red), and the ground truth (green).

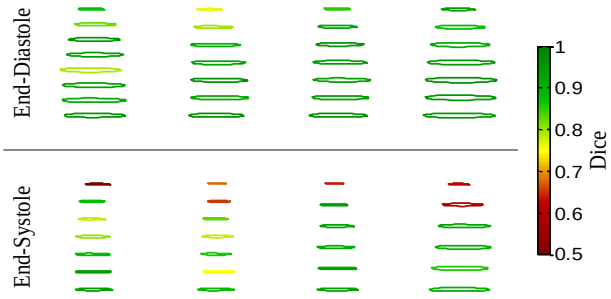


Figure 3: Examples of the estimated 3D segmentation and corresponding evaluation.

References

[1] A. Andreopoulos and J. K. Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Medical Image Analysis*, 12(3):335–357, 2008.

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[3] T. Heimann and H. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical image analysis*, 13(4):543–563, 2009.

[4] W. Gregory Hundley et al. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. *Journal of the American College of Cardiology*, 55(23):2614–2662, 2010.

[5] C. Petitjean and J. Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical image analysis*, 15(2):169–184, 2011.

[6] M. Rogers and J. Graham. Robust active shape model search. In *Computer Vision–ECCV 2002*, pages 517–530. Springer, 2006.

[7] C. Santiago, J.C. Nascimento, and J.S. Marques. A new ASM framework for left ventricle segmentation exploring slice variability in cardiac MRI volumes. *Neural Computing and Applications*, pages 1–12, 2016.

A System for the Analysis of Dermoscopy Images Using Weak Annotations

Catarina Barata[1]
ana.c.fidalgo.barata@ist.utl.pt
M. Emre Celebi[2]
ecelebi@uca.edu
Jorge Marques[1]
jsm@isr.tecnico.ulisboa.pt

[1] Institute for Systems and Robotics
Instituto Superior Técnico
Lisboa, Portugal

[2] Department of Computer Science
University of Central Arkansas
Arkansas, USA

Abstract

This paper proposes a two-step approach for the analysis of dermoscopy images. In the first step, we detected dermoscopic criteria (structures and colors), which are used by dermatologists in their medical analysis. In the second step, this information is used to automatically diagnose skin cancer.

The extraction of dermoscopic criteria from skin lesions is a challenging task because the amount of detailed annotated images is scarce. We solve this task by using a probabilistic model (topic model) learned from weakly annotated data. This approach overcomes the need for completely annotated datasets, only requiring text labels. The second step uses the detected criteria to train a Random Forest classifier. The system achieves a good classification score: sensitivity of 85.8% and a specificity of 71.1%. Nonetheless, the main advantage of this system with respect to others is its ability to justify the decision based on medical criteria.

1 Introduction

Dermoscopy is a popular modality used by dermatologists to obtain magnified images of skin lesions. Over the last years, there has been an increasing interest in the development of automatic systems for the analysis of these images in order to detect skin cancer, namely melanoma. Most systems follow a standard classification framework: i) lesion segmentation; ii) feature extraction, and iii) classification. Although the performance of these systems is good, dermatologists do not trust these systems because: i) they are black boxes and the extracted features have no medical meaning and ii) the number of wrong decisions is still too high.

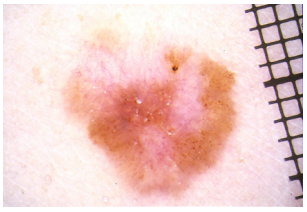
This paper aims at extracting features similar to those adopted by dermatologists in their daily procedures. Two examples are the ABCD rule [9] and the 7 point checklist [1] methods that identify specific properties inside the lesion region. These can be either relevant structures such as pigment network, dots/globules, streaks, and blue whitish veil, or specific colors, such as the six different colors considered in the ABCD rule (dark and light brown, black, white, red, and blue). The presence of high number of colors and structures in a lesion is interpreted as a malignancy cue.

The automatic extraction of such criteria is challenging. There are many attempts to solve this problem but most works only focus on a single criterion [8] and use private databases specific for that criterion. To the best of our knowledge, it is not possible to find a public database that comprises segmented images of the medical criteria.

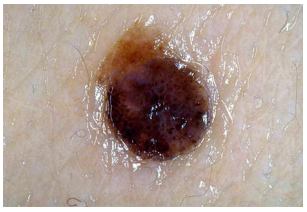
To overcome this limitation, we solve the criteria detection problem using weakly annotated data, as exemplified in Figure 1. Instead of using segmented images for each of the criterion we train a probabilistic model using text information, which states if the criterion is present or not in the image (without specifying its location). This can be done using a generative model of the image based on latent variables. The model used in this work is called Correspondence Latent Dirichelet Allocation (corr-LDA) [4, 5]. Although this model has been used with success in image annotation tasks performed in large databases, it was only recently used in the context of dermoscopy [3].

Figure 2 shows a typical input image and the output of the proposed system displaying the detected colors and structures. After extracting the medical features from the image, the system trains a support vector machine to classify new images as melanoma or benign.

This work was partially funded with grant SFRH/ BD/84658/2012 and by the FCT project [UID/EEA/50009/2013] .



Colors: Dark brown, light brown, red, and white.
Structures: Dot and regression. areas.
Diagnosis: Melanoma.



Colors: Dark brown, black, light brown, and blue.
Structures: Pigment network, dots, and blue-whitish veil.
Diagnosis: Benign.

Figure 1: Images and annotations performed by dermatologists [2].

2 Proposed System

We wish to automatically identify the presence and location of several medical criteria organized into two groups: colors (dark and light brown, black, white, red, and blue) and structures (pigment network, dots, blue whitish veil, and regression areas.)

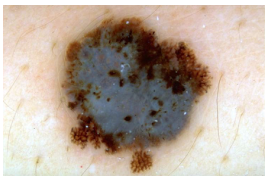
The training set consists of 804 images from EDRA database [2], each of them weakly annotated by a group of experts, *i.e.*, for each image there is a set of binary labels stating whether each criterion is present (see examples in Figure 1). On a first stage, each image is split into a tentative set if homogeneous regions, regarding color and texture. This is accomplished using the superpixel algorithm proposed in [6]. The n-th region is then characterized by a set of image features r_n (color and texture).

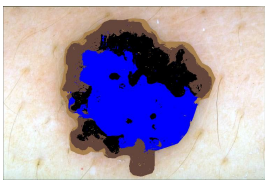
Since we want to locate each medical criteria in the image, we assume that there is a local label associated to each region. However, the information provided by the experts is a global label $w \in \{0,1\}$ valid for the entire image. Therefore, we need to find the relationship between local (region) labels and the global (image) one.

A strategy to address the previous problem consists of defining a joint probability distribution, $p(\mathbf{r},w)$, of the observed region features $\mathbf{r} = \{r_1,\ldots,r_N\}$ and image label w. One possible method to estimate this probability is the corr-LDA generative model, which uses latent variables, called topics, to increase its flexibility [4]. For each region n, the model performs three generative sequential processes, *i.e.*, it generates three variables. First, the model generates a topic, z_n, associated with the region, using a multinomial distribution. Second, it generates a feature vecto,r r_n, conditioned on the topic, using a parameterized distribution, which is defined by the user. Third, it generates the local label associated with the region w_n, which also depends on the topic z_n, using a multinomial distribution. This local label also depends on the topic z_n. After performing the generative process N times, the image label is obtained by randomly selecting one of the regions and copying the local label.

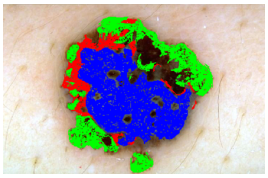
The parameters of the distributions used in the three generative steps described above have to be estimated from the training data. A maximum likelihood estimation of the model parameters is unfeasible because it is not possible to analytically compute the likelihood function. The estimation is accomplished by resorting to variational methods, namely using a variational Expectation-Maximization algorithm (see [5] for details).

Given a new image, we apply the estimated corr-LDA model and assign the most probable label to each region. This provides a segmentation of the image according to the medical criteria. The next step concerns the estimation of a global label from the model information. We train a Random Forest classifier to predict the presence of each criteria. The inputs of the classifier are the probability distribution of each label given all the





Automatic labels: Dark brown black, and blue.



Automatic labels: Pigment network, dots, and blue-whitish veil.

Figure 2: Input image (1st row) and the output of the detection color (left) and structures (right) models (2nd row). The color scheme is the following. Right image: the color scheme is green for pigment network, red for dots/globules, and blue for blue whitish veil.

features extracted from the image, $p(w|\mathbf{r})$ [3].

After performing the detection of the medical criteria, our second goal is to use them to predict a lesion diagnosis. In this stage we follow a more traditional pattern recognition approach, *i.e.*, extract features from the detected criteria (outputs of corr-LDA) and use them to train a classifier. The extracted features are: i) the presence/absence of each criteria,ii) $p(w|\mathbf{r})$, and iii) the average number of regions per topic, computed as described in [5]. Different classification algorithms have been tested and the best one was Random Forests. Thus, this is the algorithm used in this work.

3 Experimental Results

The experiments were carried on a heterogeneous dataset of 804 images (241 melanomas) from the EDRA database [2]. All of the images were analyzed by several experts during a consensus meeting. Each image is associated with a set of global text labels stating which are the observed criteria. The training and test of the annotation and classification blocks were performed using a 10-fold nested cross-validation procedure.

Since color and structures co-occur, two corr-LDA models were trained, one for color and another one for structures. This allows the assignment of two different labels to the same region. In the case of the color model, the features r_n used to describe the regions are the mean HSV values, while in the case of the structures model the features are the mean HSV values and the texture features: contrast and anisotropy.

Figure 2 shows the output of the criteria detection block for both color and texture. The model is able to correctly predict all the color and structure labels associated with the image. It also provides a segmentation of the image according to the different criteria. Although we do not have ground-truth segmentation for the image (recall that the model was trained using text labels only), the segmentation proposed by the model seems to provide a correct interpretation of the image. Nonetheless, it would be interesting to be able to validate the segmentations comparing with the medical performance. However, this is far from being simple.

Tables 1 and 2 and show the performance of label assignment. The detection of each criteria is considered as a binary decision problem, characterized in terms of precision and recall. The system correctly detects most of the structures and the colors. However, it is possible to see that the performance changes according to the structure or color considered, e.g., dark brown is detected with a precision of 95.7% and recall 95.7%, while the red and white colors are detected with lowers scores. One possible explanation is the number of examples in the training set. Brown color is very common, while white and red are rare and appear only in 24 and 39 images, respectively.

Melanoma detection can be performed using color, structures or both. We trained a separate Random Forest for color and structures. Both of them can predict the presence of melanoma and give a score $s \in [0, 1]$, where 1 is malignant and 0 is benign. We also combined their outputs using late fusion by simple average of both scores [7].

The performance of melanoma detection is shown in Table 3. Color and structures achieve comparable performances (structures perform slightly better). The combination of both criteria improves sensitivity, which is the most important metric (probability of correct detection if the lesion is a melanoma).

Table 1: Results for structure detection.

Structures	Precision	Recall
Pigment Network	78.5%	86.1%
Dots	72.8%	83.7%
Blue-whitish veil	82.8%	68.1%
Regression areas	63.9%	58.8%

Table 2: Results color detection.

Colors	Precision	Recall
Blue-Gray	87.6%	94.2%
Dark-Brown	95.7%	95.7%
Light-Brown	89.1%	92.7%
Black	81.5%	88.8%
Red	79.3%	74.2%
White	63.6%	93.3%

Table 3: Results for melanoma diagnosis using Random Forests.

Criteria	Sensitivity	Specificity
Structures	80.9%	74.8%
Colors	80.1%	71.9%
Combined	85.8%	71.1%

4 Conclusions

This paper proposes a system that extracts medical information from the image: it provides text labels of medical criteria and their location inside the lesion. This is achieved by training the system with weakly annotated images, which means that the training set is annotated by experts with text labels but no information is provided regarding their location within the image. Furthermore, the system uses this information to provide a diagnosis of the lesion as benign or melanoma. This means that dermatologists receive an automatic decision and the medical information that justifies it. To the best of our knowledge, this is the first system that provides this information.

References

[1] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and El. Delfino. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Archives of Dermatology*, 134:1563–1570, 1998.

[2] G. Argenziano, H P. Soyer, V. De Giorgi, D. Piccolo, P. Carli, M. Delfino, A. Ferrari, V. Hofmann-Wellenhog, D. Massi, G. Mazzocchetti, M. Scalvenzi, and I H. Wolf. *Interactive Atlas of Dermoscopy*. EDRA Medical Publishing & New Media, 2000.

[3] C. Barata, M. E. Celebi, J. S. Marques, and J. Rozeira. Clinically inspired analysis of dermoscopy images using a generative model. *accepted for publication in Computer Vision and Image Understanding*.

[4] D.M. Blei and M.I. Jordan. Modeling annotated data. In *26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 127–134. ACM, 2003.

[5] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[6] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2): 167–181, 2004.

[7] J. Kittler, M. Hatef, R.P. W. Duin, and J. Matas. On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence*, 20(3):226–239, 1998.

[8] K. Korotkov and R.l Garcia. Computerized analysis of pigmented skin lesions: a review. *Artificial intelligence in medicine*, 56(2):69–90, 2012.

[9] W. Stolz, A. Riemann, and A B. Cognetta. ABCD rule of dermatoscopy: a new practical method for early recognition of malignant melanoma. *European Journal of Dermatology*, 4:521–527, 1994.

Irregularity Detection in ECG signal using a semi-fiducial method

João Carvalho
joao.carvalho@ua.pt
Armando J. Pinho
ap@ua.pt
Susana Brás
susana.bras@ua.pt

IEETA
University of Aveiro
Aveiro, Portugal Portugal

Abstract

Irregularity detection is important for several ECG applications, both in the field of health or even other subjects like biometric identification. In this paper, we have proposed a method based in the *Hamming Distance* for finding irregularities on an ECG signal, that outperforms an alternative proposed in a paper by Keogh *et al*. in [6], at least for this specific dataset. We start by explaining the preprocessing steps required to work with ECG signals using non-fiducial methods and make a comparison of the results obtained by both methods.

1 Introduction

It is supposed that, at rest, the ECG signal from a complete cardiac cycle is similar to the previous and to the next cycle. However, due to external or internal interferences, this may not be true [1, 3, 8].

Developing an algorithm to identify where those interferences occur may be of great interest for biometry identification, as well as other applications, as it may allow to incorporate that algorithm into a decision support system, where parts of the signal that contain irregularities may be treated differently than regular signal.

2 Overview

In [6], the authors describe a method for finding noise using compression tools. They do this by computing the dissimilarity distance of a given segment of an ECG against the whole ECG. The *Compression-based Dissimilarity Measure* (CDM) between two string x and y [6] is defined as:

$$\mathrm{CDM}(x,y) = \frac{C(xy)}{C(x)+C(y)}, \tag{1}$$

where xy represents the strings x and y concatenated.

Two important facts about the CDM are: **(a)** it is close to one when x and y are not related; **(b)** if x and y are related, as strongly related they are, the lower the $\mathrm{CDM}(x,y)$ is, but it never reaches zero.

In short, what their method does is to "simply measure how well a small local section can match the global sequence" [6]. It is easy to see how this can be useful to find irregularities in the signal (assuming they are present only in some small portions of the signal).

Since the NRC has been shown to work very well on ECG signals [3] and it also respects both **(a)** and **(b)**, we did an implementation using it as a replacement for the CDM defined by Keogh *et al*. Finally, we compared that approach with our proposed method, based on the *Hamming Distance*.

2.1 Database

Despite the fact the proposed algorithm aims at the detection of ECG noise in real signals, we had to manage a control set of ECG signals to test and validate the algorithm. These were already acquired on a previous study.

The database is composed by 67 hours of ECG signal from which 19 hours (7 signals with 2h40 each, approximately) correspond to a clean ECG signal collected using a simulator with artificial/synthetic noise added. To gather the values from the simulator, we used VitalJacket [4], which provides an ECG signal with a sampling frequency of 500 Hz. The collecting protocol consisted in gradually increasing and decreasing the heart rate with 20 minutes interval. In total, **the signal is composed by 8 heart rate steps in the following order: 60, 80, 100, 120, 140, 120, 100, 80 beats per minute**. The noise values were randomly generated using a uniform discrete distribution between the minimum and maximum values (117 and 159, respectively) of the collected signal. This allowed us to have a controlled signal with noise in specific zones, to test the algorithm's behavior in different situations.

To each signal was then added some noise in specific zones, as we will see in some examples.

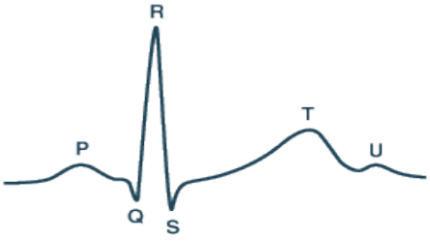


Figure 1: P, Q, R, S, T, U peaks of an ideal ECG heartbeat (from [2]).

3 ECG Signal Preprocessing

The ECG signal is not periodic, but it is highly repetitive [1]. There are three major components of a complete heartbeat captured by an ECG signal: the *P* wave, the *QRS* complex and the *T* wave. The points *P*, *Q*, *R*, *S*, *T* and *U* (which is less used) are called *points of interest*, also known as *fiducial points*, of the ECG. An example can be seen in Fig. 1.

3.1 R-peak detection

The development of a robust automatic *R-peak* detector is essential, but it is still a challenging task, due to irregular heart rates, various amplitude levels and shapes of *QRS* morphologies, as well as all kinds of noise and artifacts. [5]

We have decided to use a *partially fiducial* method for segmenting the ECG signal and, since this was not the major focus of the work, we used a preexisting implementation to detect *R-peaks*, based on [5]. This method detects the *R-peak* by calculating the average point between the *Q* and *S* peaks (the *QRS complex*) – this may not give the real local maximum of the *R-peak*, but it produces a very close point. Some validations were done against *R-peaks* detection performed by humans in order to validate this step.

The process used for detecting the *QRS* complexes is somewhat similar to the one described in [5]. It uses some bandpass filtering and differentiation operations used to enhance *QRS* complexes and to reduce out-of-band noise. A nonlinear transformation based on energy thresholding, Shannon energy computation, and smoothing processes is used to obtain a positive-valued feature signal which includes large candidate peaks corresponding to the *QRS* complex regions.

3.2 Quantization

In order to convert the real-valued ECG signal into a symbolic time series, the first step we had to perform was to reduce its dimensionality. This was achieved by using a modification of the Piecewise Aggregate Approximation, PAA, method [7].

Since, in our case, the *R-peaks* of the ECG signal were already detected, we used this to our advantage and, instead of splitting the complete

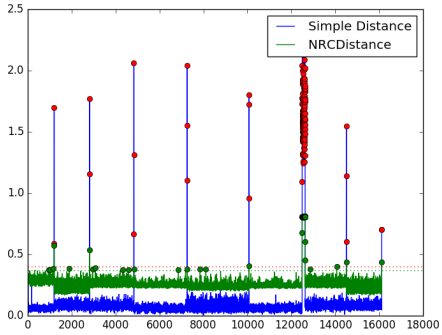


Figure 2: The transition between 120 and 100 beats per minute was replaced by 90 seconds of noise (signal number 5).

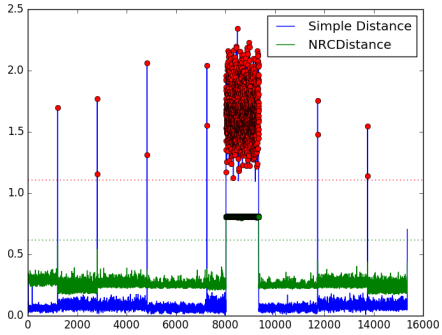


Figure 3: In the 140 heart rate step, 15 minutes of noise were introduced (signal number 3).

signal into w dimensions, as the original method suggests, we applied it only inside of each *R-R-interval* (intervals between consecutive *R-peaks*) individually, with the values inside each of those intervals previously normalized. This means that, for example, if one heartbeat takes more time than another, then the real value obtained by PAA may correspond to more real values on the original signal. On our case, the number of samples, w, chosen for each heartbeat was 200. This means that each heartbeat on the ECG, on this phase, was represented by 200 real values (instead of the original 1000 real values per second).

After completing this process for all the heartbeats of a signal, since the purpose is to have a symbolic representation of the series – not a real valued one, the *Symbolic Aggregate approXimation* (SAX) was applied to each heartbeat PAA series individually.

From this explanation, it is already implicit that one parameter used by this method as input is the *alphabet size* (the number of different symbols allowed as output), that we want to use. From experiments using a different database, we realized a choice of an alphabet size of 6 is appropriate for most of our applications and, therefore, this was the alphabet size we used for this tests as well. Using the process described, each complete heartbeat is outputted as a 200 length string. We refer to a string like that as a *word* or *SAX-word*.

3.3 Proposed Method

The proposed method is very straight forward, which is why we called it the *Simple Distance method* (in fact, it basically consists on the *Hamming Distance* applied to consecutive *SAX-words*).

The idea of our approach is to store all the n *words* (or *SAX-words*) of an ECG on a size n array and compute the $n-1$ distances between those consecutive n words. Since we want to have a sample of size n, an interpolation from size $n-1$ to size n should be performed. After that, some decisions can be made using the values obtained for that metric.

3.4 Results and Future Work

From our experiments, we found that a threshold which produced more consistent results for both metrics was $\bar{x} - 2\sigma$ (average value minus two times the standard deviation). We were able to do it by experimentation because we knew the zones where noise was supposed to be found beforehand.

In Fig. 2, it is possible to see that the *Simple Distance* measure detects the areas where random noise was inserted very precisely, while the NRC detects a lot of false positives. In Fig. 3 the *NRC* is not able to detect any noise at all, however, the *Simple Distance* can detect both the zones where the heart beat rate was changed, which, even tough it is not noise, may be considered a point of interest, depending on the application.

Even though the threshold choice worked properly for this dataset, it should not be static and, therefore, some future work should be done in order to adjust it in a dynamic way. We also plan on making further tests using different datasets, as well as possible changes to the method itself, which is still very basic.

4 Acknowledgments

This work was partially supported by national funds through the FCT - Foundation for Science and Technology, and by european funds through FEDER, under the COMPETE 2020 and Portugal 2020 programs, in the context of the projects UID/CEC/00127/2013 and PTDC/EEI-SII/6608/201

References

[1] Foteini Agrafioti and Dimitrios Hatzinakos. ECG biometric analysis in cardiac irregularity conditions. *Signal, Image Video Process.*, 3(4):329–343, sep 2008. ISSN 1863-1703. doi: 10.1007/s11760-008-0073-4. URL http://link.springer.com/10.1007/s11760-008-0073-4.

[2] M Bassiouni and W Khalefa. A study on the Intelligent Techniques of the ECG-based Biometric Systems. *Recent Adv. Electr. Eng.*, 2015. URL http://www.inase.org/library/2015/crete/COCI.pdf{\#}page=26.

[3] Susana Brás and Armando J Pinho. ECG biometric identification: A compression based approach. In *2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 5838–5841, aug 2015. doi: 10.1109/EMBC.2015.7319719. URL http://www.ncbi.nlm.nih.gov/pubmed/26737619.

[4] João P. Silva Cunha, Bernardo Cunha, António Sousa Pereira, William Xavier, Nuno Ferreira, and Luis Meireles. Vital-Jacket: A wearable wireless vital signs monitor for patients' mobility in cardiology and sports. In *2010 4th Int. Conf. Pervasive Comput. Technol. Healthc.*, pages 1–2. IEEE, 2010. doi: 10.4108/ICST.PERVASIVEHEALTH2010.8991.

[5] P. Kathirvel, M. Sabarimalai Manikandan, S. R. M. Prasanna, and K. P. Soman. An Efficient R-peak Detection Based on New Nonlinear Transformation and First-Order Gaussian Differentiator. *Cardiovasc. Eng. Technol.*, 2(4):408–425, oct 2011. ISSN 1869-408X. doi: 10.1007/s13239-011-0065-3. URL http://link.springer.com/article/10.1007/s13239-011-0065-3/fulltext.html.

[6] E Keogh, S Lonardi, and CA Ratanamahatana. Towards parameter-free data mining. In *KDD '04 Proc. tenth ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pages 206–215, 2004. URL http://dl.acm.org/citation.cfm?id=1014077.

[7] J Lin, E Keogh, S Lonardi, and B Chiu. A symbolic representation of time series, with implications for streaming algorithms. In *DMKD '03 Proc. 8th ACM SIGMOD Work. Res. issues data Min. Knowl. Discov.*, pages 2–11, 2003. URL http://dl.acm.org/citation.cfm?id=882086.

[8] 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. *IEEE Trans. Inf. Forensics Secur.*, 7(6):1812–1824, dec 2012. ISSN 1556-6013. doi: 10.1109/TIFS.2012.2215324. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6287579.

Machine Learning with Word Embeddings applied to Biomedical Concept Disambiguation

Rui Antunes
ruiantunes@ua.pt

Sérgio Matos
aleixomatos@ua.pt

IEETA
University of Aveiro

Aveiro, Portugal
http://bioinformatics.ua.pt

Abstract

Artificial Intelligence (AI) has grown in the last years and it has many applications. Natural Language Processing is one of the AI tasks, which has the objective to endow the machines the capability of understanding human language. This is an important process due to the amount of information stored in textual form. There is a growing need for automatic extraction of knowledge, and NLP comes in this direction helping in tasks such as information extraction and information retrieval. Word sense disambiguation is an important NLP subtask, which is responsible for assigning the proper concept to an ambiguous word or term.

In this paper, we present results obtained from applying supervised machine learning algorithms with local features, and word embeddings as global features extracted from Wikipedia and PubMed knowledge sources. These results indicate that word embedding features are informative and may improve the biomedical word disambiguation accuracy.

1 Introduction

Large volumes of biomedical data are produced every day, and this is accompanied by an also increasing amount of textual data, mostly in the form of scientific publications. In order to efficiently treat and interpret these data it is necessary to create tools that automatically do this job, reducing the human efforts. This led to the application of text mining methods for extracting information from the literature and linking that to repositories of biomedical data. For instance, the work in [1] describes a framework for biomedical concept recognition, which is a relevant task for biomedical Information Extraction (IE).

Word Sense Disambiguation (WSD), an important subtask of Natural Language Processing (NLP) [2], is a challenging task that consists of finding the correct sense of an ambiguous term. Usually, this is achieved using the surrounding context of the term. Currently, there are mainly two distinct approaches for WSD, those based on Machine Learning (ML) algorithms and the ones based on knowledge sources. The ML approaches can follow supervised, semi-supervised or unsupervised algorithms, with supervised classification approaches currently offering the best results in terms of accuracy, achieving around 94% using a Support Vector Machine (SVM) classifier [3].

Knowledge-based approaches to WSD have also attracted large interest, as these approaches are usually less dependent on training data, which may lead to better generalization when compared to supervised learning algorithms. The use of multiple knowledge bases brings benefits to the problem of concept disambiguation [4]. WordNet [5] is a large knowledge database of the English language that has been extensively applied for word sense disambiguation [2]. In the case of biomedical texts, the largest and most relevant knowledge database is the Unified Medical Language System (UMLS) [6], which offers a rich integrated metathesaurus and semantic network for this domain.

Word embeddings is a recent technique, which can be applied in NLP. It converts words from a document collection, or corpus, into vectors of real numbers. These word embeddings can be used as global features in a ML classification problem. In our case, these features were used in the disambiguation task, which showed to be almost as effective as local features. In [3], the authors present a work on supervised biomedical word sense disambiguation applied to the MSH WSD data set [7], exploring the combination of unigrams as local features and word embeddings as global features. Other approaches using word embeddings for word sense disambiguation have also been proposed by Wu et al. [8], and Taghipour and Ng [9].

In this work, we applied several machine learning methods in the MSH WSD data set in order to measure the WSD accuracies. The ML classifiers used in this experiment were the decision tree classifier, the k-nearest neighbors vote, the passive aggressive linear model, the ridge regression classifier, the Support Vector Machine (SVM) classifier. Textual data from Wikipedia and PubMed corpus were used to generate the word embeddings features to be used in the classifiers.

2 Data Set

The MSH WSD data set was automatically generated using the UMLS Metathesaurus and MEDLINE citations [7]. The data consisting of scientific abstracts, each with one ambiguous term identified and mapped to the correct sense. It contains 203 ambiguous terms with a total of 423 distinct senses. Most terms (189) have only two different meanings, 12 terms have three different meanings, and the remaining 2 terms have four and five different meanings. There are a total of 37,888 examples of ambiguity. Each term has, on average, 187 citations, that is, ambiguity cases.

3 Methods

For each ambiguous term, we applied 5-fold cross-validation to subdivide the corresponding abstracts for training and testing the model. A bag-of-words (BOW) model was used to represent the texts, with local features acquired from the context, namely unigrams and bigrams, with tf-idf weighting. In order to evaluate the impact on WSD accuracy, we also added word-embedding vectors, calculated from Wikipedia and PubMed corpora, as global features. A list of 313 stopwords obtained from the Medline repository[1] was used to filter out very frequent words in the corpus. All these tasks were implemented using the framework Scikit-learn [10], a machine-learning library for the Python programming language. Word embedding models were obtained with the gensim framework [11].

3.1 Machine Learning Methods

In order to obtain the highest accuracy, several machine learning classifiers were compared: decision tree, k-nearest neighbor, passive aggressive linear model, ridge regression, SVM.

3.2 Feature Combination

The local features used were unigrams and bigrams, and the global features used were the word embeddings. We tested different feature combinations in order to understand which combination produced the best results. Local features were scaled using the term frequency – inverse document frequency (tf-idf) scheme.

Table 1: WSD accuracies with only global features (Wikipedia model vs PubMed model). Results shown are the average across five folds. DT: Decision Tree; kNN: k-Nearest Neighbor (k=5); PA: Passive Aggressive linear model; RR: Ridge Regression; SGD: linear Support Vector Machine with Stochastic Gradient Descent; SVC: Support Vector Classification.

	Model of word embeddings from	
	Wikipedia	PubMed
DT	0.817	0.849
kNN	0.896	0.918
PA	0.893	0.928
RR	0.905	0.910
SGD	0.874	0.916
SVC	0.912	0.924

[1] https://mbr.nlm.nih.gov/Download/2009/WordCounts/wrd_stop

Table 2: Accuracies using distinct features combinations. Results shown are the average across five folds. U: Unigrams; B: Bigrams; WE: Word Embeddings with PubMed model; DT: Decision Tree; kNN: k-Nearest Neighbor (k=5), PA: Passive Aggressive linear model; RR: Ridge Regression classifier; SGD: linear Support Vector Machine with Stochastic Gradient Descent; SVC: Support Vector Classification.

	Local features			Local and global features
	U	B	U+B	U+B+WE
DT	0.903	0.862	0.901	0.908
kNN	0.913	0.918	0.924	0.919
PA	0.950	0.938	0.949	0.934
RR	0.942	0.922	0.940	0.939
SGD	0.947	0.931	0.946	0.919
SVC	0.948	0.932	0.948	0.938

3.3 Word Embeddings

Two distinct models of word embeddings were calculated, from Wikipedia and PubMed articles respectively. Wikipedia is range-wide, having no specific domain. The full Wikipedia dump, obtained in September 2015, was used, amounting to approximately four million articles and containing about two million distinct words. PubMed, on the other hand, is specific to biomedical domain. Around six million abstracts corresponding to the years 2010 to 2015 were used, containing around 400 thousand distinct words. Both models were trained with a window of five words and for a feature vector of size 100. Each abstract (instance) was represented by the weighted average of the embedding vectors for all the words in the abstract, with the tf-idf value of each word used as weight.

4 Results

First, we compare the two distinct models of word embeddings as unique features of the classification problem in order to find the best model to use (Table 1). As expected, the model from the domain specific PubMed corpus outperformed the more general model created from Wikipedia articles. Nevertheless, the results obtained with the latter indicate that even features extracted from general corpora may contribute to these methods.

The results in Table 2 show that the state-of-the-art results for this problem can be reproduced using simple word-based features. It is also noticeable that bigram features contribute only slightly to the results, and unigram features alone achieve almost as good if not better results than the combination of unigram and bigrams. Also, comparing these results with Table 1, one can observe that word embedding features alone, which in this study represent vectors of 100 features, allow obtaining results that are very close to the best results obtained with unigram features.

On the other hand, the combination of word embedding features with unigrams and bigrams did not improve results. In our experiments the highest accuracy, 95.0%, was obtained with unigram features alone, using the passive-aggressive linear classifier.

5 Discussion

As expected, the word embeddings model from PubMed outperformed the Wikipedia model, since PubMed is specific to the biomedical domain. In our experiments, the best accuracy was attained with simple unigram features, and adding bigram features only improved the results in the case of the kNN classifier.

Disambiguation using word-embedding features alone proved very positive for this data set, even when using a small portion of the full MEDLINE database, which contains around 22 million abstracts. However, the combination of these and simple word-based features lowered the classification accuracy. Jimeno-Yepes [3] achieved an accuracy of 95.97%, in this same data set, with the combination of unigrams and word embeddings from the full MEDLINE. In future work, we will extend this analysis and investigate different strategies for integrating the word embedding features in this classification problem.

References

[1] D. Campos, S. Matos and J. L. Oliveira. A modular framework for biomedical concept recognition. *BMC Bioinformatics*, 14:281, 2013.

[2] R. Navigli. Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, 41(2):10, 2009.

[3] A. J. Jimeno-Yepes. Higher order features and recurrent neural networks based on Long-Short Term Memory nodes in supervised biomedical word sense disambiguation. *arXiv: 1604.02506v1 [cs.CL]*, 2016.

[4] C. T. Tsai and D. Roth. Concept Grounding to Multiple Knowledge Bases via Indirect Supervision. *Transactions of the Association for Computational Linguistics*, 4:141–154, 2016.

[5] C. Fellbaum. WordNet: An electronic lexical database. *Cambridge: MIT Press*, 1998.

[6] O. Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32:267–270, 2004.

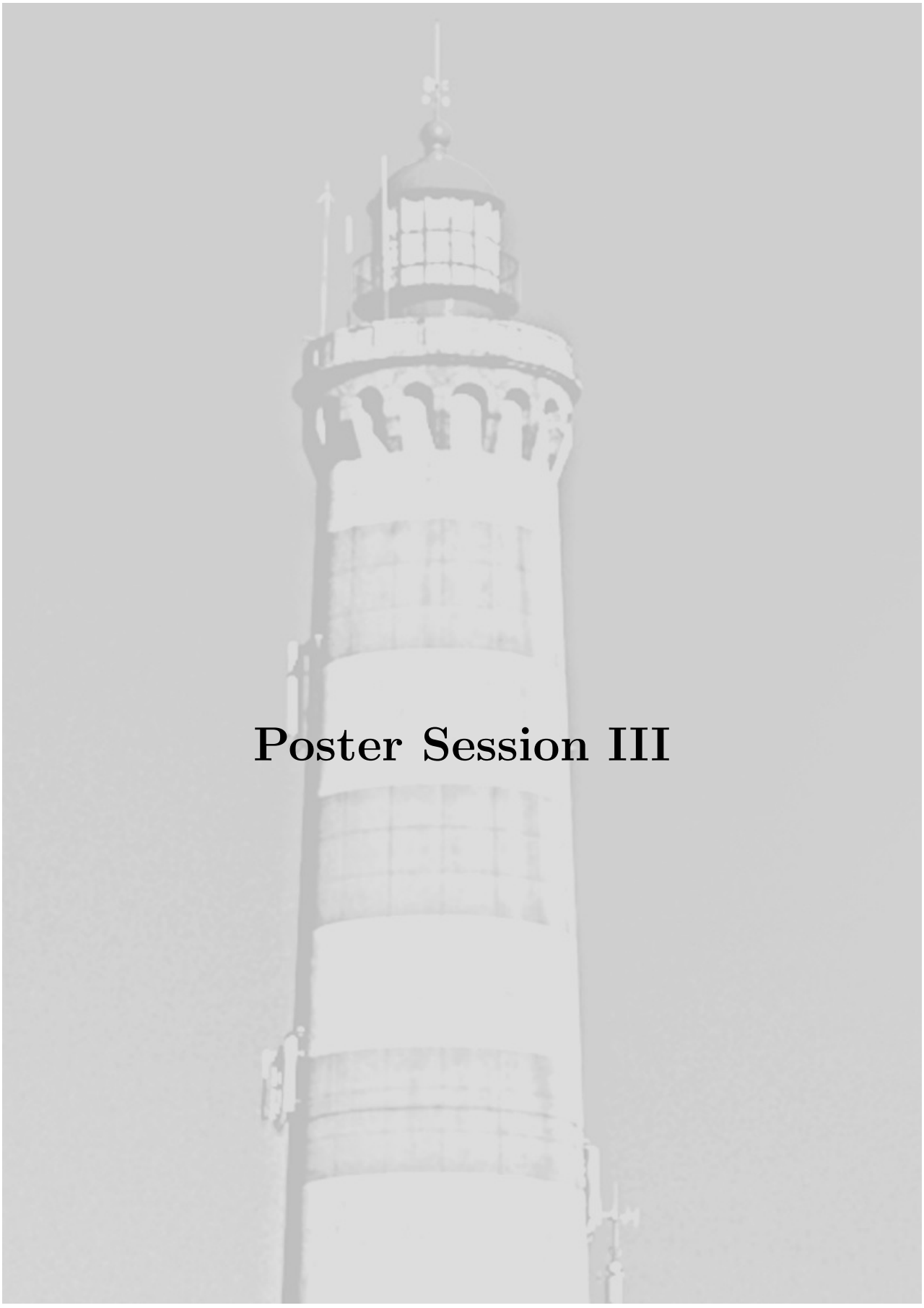
[7] A. J. Jimeno-Yepes, B. T. McInnes and A. R. Aronson. Exploiting MeSH indexing in MEDLINE to generate a data set for word sense disambiguation. *BMC Bioinformatics*, 12:223, 2011.

[8] Y. Wu, J. Xu, Y. Zhang and H. Xu. Clinical Abbreviation Disambiguation Using Neural Word Embeddings. *ACL-IJCNLP*, pages 171–176, 2015.

[9] K. Taghipour and H. T. Ng. Semi-Supervised Word Sense Disambiguation Using Word Embeddings in General and Specific Domains. *Proceedings of NAACL HLT*, pages 314–323, 2015.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[11] R. Rehurek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, 2010.



Poster Session III

Intrinsic Page Hinkley Test (iPHT)

Raquel Sebastião
raquel.sebastiao@ua.pt
José Maria Fernandes
jfernan@ua.pt

IEETA, DETI
University of Aveiro
Campus Universitário de Santiago
3810-193 Aveiro, Portugal

Abstract

As a data stream is gathered over time and for large periods, changes in data are expected. In dynamic scenarios it is of utmost importance to perform change detection tests, in order to detect changes as soon as possible. Along with this, the false and the missed detections must be minimal, establishing a trade-off between the robustness to false detections and the sensitivity to true changes. We propose the Intrinsic Page Hinkley Test (iPHT), which enhances the PHT by using intrinsic functions from the Empirical Mode Decomposition (EMD). The PHT is a sequential analysis approach for detecting changes in data and the EMD is a time-frequency analysis technique designed to work well for nonstationary and nonlinear data. Therefore, the proposed iPHT can be applied without any prior hypothesis on the data as both PHT and EMD can be applied over any arbitrary data stream without any constraints.

1 Proposed Strategy

The proposed strategy relies on the enhancement of a well known test used for detecting changes by reducing the number of user-defined parameters of this test. The Page Hinkley Test (PHT) requires an input threshold and an input parameter, which are defined by the user according to the data properties and application purposes. By supporting the PHT with the Empirical Mode Decomposition (EMD) we can replace one input parameter by an Intrinsic Mode Function (IMF). The EMD is the fundamental part of the Hilbert–Huang transform, proposed by [2, 5]. It can successfully handle non stationary and nonlinear data and allows maintaining the data handling on the time domain. Indeed, although EMD is a time-frequency analysis method, it does not imply time to frequency transformations. Moreover, unlikely other signal processing decomposition methods, EMD makes no assumptions on the incoming data. This means that our change detection method can be applied without any prior hypothesis on the data as both PHT and EMD can be applied over any arbitrary data stream without any constraints.

1.1 The page-Hinkley Test (PHT)

The Page Hinkley Test (PHT) [3, 4, 6] is a sequential analysis technique typically used for monitoring change detection in the average of a Gaussian signal [1]. This test considers a cumulative variable defined as the accumulated difference between the observed values and their mean until the current moment. The two-sided PHT detects both increases and decreases in the mean of a sequence, running two tests in parallel and reporting a change whenever one of the two monitored statistics is greater than a giver threshold.

1.2 The Empirical Mode Decomposition (EMD)

The Empirical Mode Decomposition (EMD) is the fundamental part of the Hilbert–Huang transform, which was proposed by [2, 5], and is designed to work well for data that is non stationary and nonlinear. The EMD does not make assumptions on the incoming data. There are several free implementations of EMD available: in this work we used the one developed by Alan Tan [1].

The EMD consists of successively decomposing the original data $x(t)$ into Intrinsic Mode Functions (IMF) $c_i(t), i = 1,\ldots,n$ and into the monotonic residual $r(t)$. Once the first IMF is removed from the original data, the procedure is successively applied to the residual. This process will decompose the original data into the highest frequency component (c_1) to the lowest frequency component (c_n), until the residual $r(t)$ is a monotonic function from which no more IMF can be extracted: $x(t) = \sum_{i=1}^{n} c_i(t) + r(t)$, where n is the number of IMF. An IMF is an oscillating wave [5], which can have variable amplitude and frequency along the time axis and it must satisfy the two following requirements:

1. The number of extrema and the number of zero-crossings differs, at most, by one.
2. The local average value of the upper and the lower envelope is zero.

The process of extracting an IMF is called sifting and it is repeated until the above two conditions of an IMF are satisfied.

[1] available at http://www.mathworks.com/matlabcentral/fileexchange/19681-hilbert-huang-transform (accessed in Sep 8th 2016)

1.3 The Intrinsic Page Hinkley Test (iPHT)

The EMD allows an effective time-frequency analysis that can be used to filter a data stream. Indeed, the lower IMF levels are related with higher frequency of the data and higher IMF levels describe the data baseline wander. Therefore, instead of removing the high-frequency noise of the data by reconstruction without the lowers IMF levels, one are interested in using the lower IMF levels to establish the magnitude of the changes that are allowed. In this context the PHT is supported through the EMD, by replacing the δ parameter with an IMF, which corresponds to the magnitude of changes that are allowed without triggering an alarm. Hence we are loosing one user-defined parameter of the PHT by using a component that will depend on the data itself. The iPHT tests are the following:

For increase cases:	For decrease cases:
$iU_0 = 0$	$iL_0 = 0$
$iU_T = (iU_{T-1} + x_T - \bar{x}_T - IMF_T)$	$iL_T = (iL_{T-1} + x_T - \bar{x}_T + IMF_T)$
($\bar{x}_T$ is the mean until the current sample)	
$m_T = min(iU_t, t = 1\ldots T)$	$M_T = max(iL_t, t = 1\ldots T)$
$iPH_U = iU_T - m_T$	$iPH_L = M_T - iL_T$

At each observation, the two PH statistics (iPH_U and iPH_L) are monitored and a change is reported whenever one of them rises above a given threshold λ. The threshold λ depends on the admissible false alarm rate: increasing λ will entail fewer false alarms, but might miss or delay some changes.

To choose the order of the IMF derived from the EMD that should be used to replace the PHT user-defined parameter δ, the relation between the several IMF orders and the original data must be take into consideration. The information shared between the original data and the IMF levels decrease with the level: the first IMF shares more information with the original data. As the δ parameter corresponds to the magnitude of the changes that are allowed, the original data and the IMF level must be positive related. On the other side, they must not share too much information. Therefore, the order of the IMF used to control the magnitude of changes that are allowed will be studied, in the section 2, throughout the evaluation of the covariance and the Spearman correlation between the data and several IMF orders.

2 Results and Discussion

In this section, the proposed iPTH is evaluated under different evolving scenarios, using artificial data. The data sets were generated according to normal distributions, varying the mean parameter (and with $\sigma = 1$). Each data stream consists of 2 parts, each of with size $N = 5000$. Different changes were simulated by varying among 3 levels of magnitude and 3 rates of change, obtaining a total of 9 types of changes. For each type of changes, 10 different data streams were generated with different seeds. More details on the artificial data design can be found in [7]. Moreover, using similar data, a training and a test set were produced. A real data set obtained from an industrial context was also used to evaluate the iPHT (more details on the dataset in [7]).

2.1 Artificial Data

Choosing the order of the Intrinsic Mode Function

This first experiment was designed to assess the relation between the original data and several IMF levels: the covariance and the Spearman correlation were evaluated on the training set.

Figure 1 shows the covariance between the original data and the correspondent five first IMF levels (average with error bars is reported for 10 runs on data generated with different seeds). The Spearman correlation presented a similar behavior to the covariance (not shown due to lack of space). For both measures, the higher values are obtained for the IMF_1. When comparing the original data with the IMF_1 and with the remain IMFs, it can be observed a steeper decay in the covariance than in the Spearman correlation. Moreover, it can be noted a repeated pattern along the different changes. Therefore, and considering this results on the shared information, the second order of the IMF was chosen to replace the usually user-defined δ parameter of the PHT.

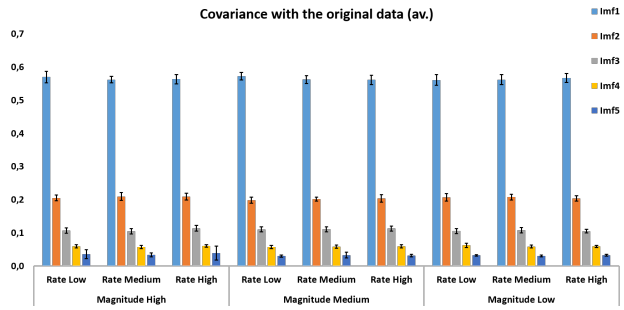


Figure 1: Covariance between the data and the five firsts IMF levels.

Study the sensibility of the algorithm input parameters

The sensibility of the λ parameter was assessed considering the total number of false alarms, the total number of missed detections and the mean detection delay time (DDT).

In this experiment, the λ parameter was varied from 100 to 1000. Independently of the λ value, the proposed iPHT detects all the changes. We also notice the absence of false alarms for λ greater than 300. As expected, the DDT increases with λ (average results over 10 runs).

Comparison with the PHT

This experiment assesses the advantage of iPHT over PHT. For that, the λ parameter was set to 500 and δ varied from 0.0001 to 1. Independently of the δ value, both the approaches detected all the changes without false alarms. The DDT of the iPHt was 733 ± 66 (average and standard deviation of 10 runs for the 9 types of changes). For the PHT, as the δ varied from 0.0001 to 1, the DDT varied from 735 ± 64 to 1818 ± 31 (average and standard deviation of 10 runs for the 9 types of changes). Thus, the proposed iPHT has the advantage of adjusting only one input parameter.

2.2 Industrial Data

The data used in this experiment is shown in figure 2 and has 6 changes. The $|delta$ was set to 2000, and λ was set to 1 (for the PHT).

The obtained results are shown in table 1. As it can be observed, the iPHT outperforms the PHT: it detects all the 6 changes with smaller DDT, although presenting 8 false alarms against 7 of the PHT.

Table 1: Detection Delay Time for PHT and iPHT, in the industrial data

Change		45000	90000	210000	255000	375000	420000	μ
DDT	PHT	350	MD	11660	10160	19140	980	8458
	iPHT	330	170	7930	2740	9170	730	3512

3 Conclusions and Further Research

The growing number of modern applications that produces evolving data underlines out the mandatory need to forget out-dated data, which does not describe the current state of the nature. We proposed the iPHT, which enhances the PHT with the EMD. The EMD can successfully handle non stationary and nonlinear data, keeping the data handling on the time domain while performing change detection tests. EMD allows an effective time-frequency analysis that can be used to filter a data stream. Indeed,

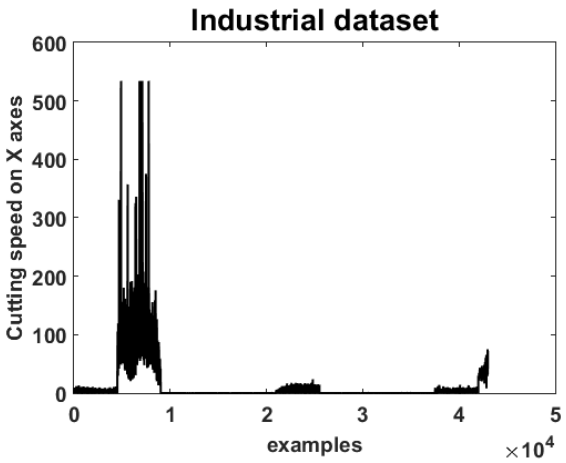


Figure 2: The cutting speed on X axes from 7 tests sequentially joined.

the lower IMF levels are related with higher frequency of the data and higher IMF levels describe the data baseline wander. Therefore, considering these properties and the amount of shared information, the results revealed that the second order IMF is suitable to replace the δ user-defined parameter of the original PHT. Further research will be devoted to a deeper evaluation of the iPHT, namely in studying the robustness against noise. Furthermore, the iPHT must be extended to a blockwise approach. Computing the intrinsic mode functions (IMF) of the EMD over sliding windows, advances a contribution to an online strategy, allowing the analysis of the data almost at the same time as it is gathered. Indeed, the short time required to evaluate the samples and the low computational complexity of the iPHT could be suitable for online processing, namely for detecting changes in internet of things and sensor networks context for human and environmental monitoring (e.g. activity levels, biomarkers indicators, pollution levels, among others).

4 Acknowledgment

This work was supported by the project VR2market (funded by the CMU Portugal program, ref. CMUP-ERI/FIA/0031/2013). The Post-Doc grant of Raquel Sebastião (BPD/UI62/6777/2015) is also acknowledged.

References

[1] H. Mouss et al. Test of page-hinckley, an approach for fault detection in an agro-alimentary production system. In *Control Conference, 2004. 5th Asian*, volume 2, pages 815 –818, july 2004.

[2] Norden E. Huang et al. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998. doi: 10.1098/rspa.1998.0193.

[3] David V. Hinkley. Inference in Two-Phase Regression. *Journal of The American Statistical Association*, 66:736–743, 1971. doi: 10.1080/01621459.1971.10482337.

[4] David V. Hinkley and Elizabeth A. Hinkley. Inference about the change-point in a sequence of binomial variables. *Biometrika*, 57 (3):477–488, 1970. doi: 10.1093/biomet/57.3.477.

[5] Norden E. Huang and S. R. Long. The mechanism for frequency downshift in nonlinear wave evolution. *Advances in Applied Mechanics*, 32:59–111, 1996. doi: 10.1016/S0065-2156(08)70076-0.

[6] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1-2): 100–115, 1954. doi: 10.1093/biomet/41.1-2.100.

[7] Raquel Sebastião. Learning from Data Streams: Synopsis and Change Detection. *PhD Thesis*, pages University of Porto, Portugal, 2014.

Pattern Recognition in Images of Counterfeited Documents

Rafael Vieira[1]
2141500@my.ipleiria.pt
Catarina Silva[1,2]
catarina@ipleiria.pt
Mário Antunes[1,3]
mario.antunes@ipleiria.pt
Ana Assis[4]
ana.assis@pj.pt

[1]School of Technology and Management, Polytechnic Institute of Leiria, Portugal
[2]Center for Informatics and Systems of the University of Coimbra, Portugal
[3]Center for Research in Advanced Computing Systems, INESC-TEC, University of Porto, Portugal
[4]Scientific Police Laboratory – Judiciary Police, Portugal

Abstract

Pattern recognition techniques are invaluable approaches to apply to forgery detection of official documents. Forgers are increasingly resorting to more sophisticated techniques to produce counterfeited documents, trying to deceive criminal polices and hamper their work. Hence, different approaches are being pursued, but seldom with real applications in real scenarios. An important challenge is the forger's *modus operandi* characterization, making it possible to obtain more information about the source of the counterfeited document.
In this paper we present a framework conceived for the Scientific Police Laboratory of the Portuguese Judiciary Police to automate counterfeit documents identification by comparing a given fraudulent document image with the images stored in a database of previously catalogued counterfeited documents.
The proposed system improves the counterfeit identification and reliefs the error prone, manual and time consuming tasks carried on by forensic experts. The framework is based on a scalable algorithm under the OpenCV framework, to compare images, match patterns and analyse textures and colours.

1 Introduction

Counterfeited documents are reproductions or imitations of the originals ones. The process of counterfeited documents identification is mostly manual and supported on expert's past experience.

The manual analysis of all the constituent elements of the questioned document is mainly based on a digital version of the original document[1,2,3] produced by using materials and printing techniques from available technologies. It is then carried out through different techniques and methodologies (physical and chemical examinations). Those elements may include printing process, watermarks, fluorescent fibers and planchettes, guilloche pattern, fluorescent and magnetic inks, optically variable inks, rainbow printing, microprinting, latent images, scrambled indicia, laser printing, photos, signatures, embossing stamps, optically variable devices, protective films, perforations, machine readable security, retro-reflective pattern, among others. This analysis provides information that may lead to the classification of the original document as genuine, false or forged.

Technical observations are based on information that may conduct to the discovery of the counterfeiting operation, i.e. associate the counterfeit with the components of its production. If a match of the counterfeited document against the database of old cases is found, the counterfeit is identified and a correlation with all the identical cases already detected in past will be successfully pursued. Otherwise, it will be created a new counterfeit number for future correlations.

The whole process of comparing a fake document with a list of previously catalogued counterfeited ones is usually made manually by the forensic experts of the Scientific Police Laboratory. Having in mind that the catalogue of documents, even for a specific document type, is potentially overwhelming, the time involved in such manual analysis may thus be prohibitive and certainly inefficient for a fast criminal investigation response. Hence, an information system based on image detection algorithms that could automate, or semi-automate such process could bring numerous advantages.

In this paper we propose a methodology to automate the comparison of a counterfeit document with an existing database of already classified counterfeit documents. The main goal is to implement an algorithm that ranks the level of similitude of an image of the questioned document being compared and thus to discard automatically those documents with less or no similitude.

In any case, the human should always remain in the loop. As such, manual verification should always be carried out in any case. However, by discarding a set of documents with less similitude probability, forensic experts' attention may be directed to the most relevant documents.

2 Image Processing Algorithms

OpenCV is a very well-known and widely used open source library for computer vision. It has tools for digital image processing and includes a set of algorithms for pattern detection and image comparison, briefly explained below.

The Harris Corner Detection algorithm[5] was developed by Chris Harris and Mike Stephens. The underpinning mathematic model used to detect corners and edges considers window in the image and then determine the average changes of image intensity. The result obtained is achieved by shifting the window by a small number of pixels in various directions. The detection of corners in digital images is made by comparing the same area in both documents.

Lowe's[6] Scale-Invariant Feature Transform algorithm (SIFT) is another algorithm to detect corners. SIFT is meant to be invariant to image scale and rotation, that is invariant when the image is zoomed out or zoomed in.

Bay et al. proposed Speeded-Up Robust Features (SURF) as a variation of SIFT, aiming to obtain an optimized version of the image to computer vision processing[7]. Rosten and Drummond developed the Fast Algorithm for Corner Detection (FAST)[8] that may have better performance in real time applications. Rublee et al. implemented Oriented Fast and Rotated Brief algorithm (ORB)[9], which is basically a mixture of SIFT and FAST algorithms.

OpenCV framework has a wide set of interesting functionalities for pattern detection in digital images, besides the core implementation. Homography is one of such functionalities that refers to the detection of an image inside another, by a matching templates.

3 Automation of Counterfeited Documents Correlation

Figure 1 depicts the image processing algorithm and the data flow used to automate the analysis of counterfeited documents. There are three distinct tasks: 1) texture analysis; 2) comparison of image areas; 3) detection of similar imperfections in text areas.

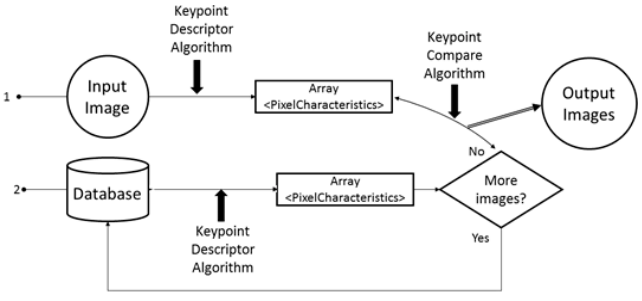


Figure 1: Image processing algorithm data flow

In (1) a given image will be processed with a "keypoint descriptor" algorithm that is part of OpenCV SURF, SIFT or ORB implementations. The output of the image processing is an array with the following information: angles of corners, edges, pixel's intensity and directions of the most pronounced intensity changes. In (2) we apply a "descriptor" algorithm to the images stored in the database that meet the same characteristics (type of document and country) of the input document. For example, if the input document is a Portuguese driver license, only documents of this type will be processed. The information retrieved for

each processed document is then compared with the array obtained in (1), using a "descriptor compare" algorithm. The output of this processing is an array with an index of similitude between the input document and each one of the stored documents that were analysed.

Texture area identification extracts the necessary parameters and is calculated by HogDescriptor texture descriptor processing algorithm, a OpenCV native algorithm.

Keypoint descriptors processing was made by SURF implementation, a non-native OpenCV algorithm. The algorithm receives an image to a keypoint descriptor for each pixel and then computes an "interestingness" function, which measures the likelihood and uniqueness of each point in another similar image. Keypoint descriptor algorithm analyzes the area around each pixel (the corners) and calculates statistical values and hashes that will be retrieved for future comparison. The algorithm continues by applying the same "keypoint descriptor" algorithm to all the other images stored on database that match the criteria (type and country) and further evaluate their level of similitude with the original input document.

3.1 Experimental Setup and Result Analysis

Our tests focused on Portuguese driver's license cards[10]. The input image is a faked stamp area, as represented in Figure 2. It was compared with images of the same type of official documents from the counterfeited images database, which in this case consists of almost 1500 entries of counterfeited driver's license cards.

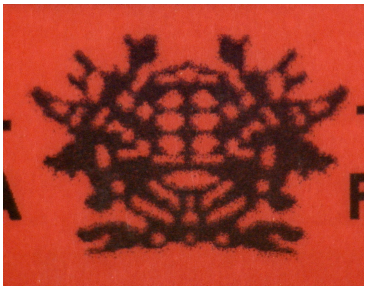


Figure 2: Input image

Using the SURF algorithm, we have calculated the closeness between the images in the dataset and the input image, by identifying the most important keypoints in each image. For each identified keypoint the respective percentage of similitude is calculated. The final score is the arithmetic average of all keypoints. Additionally, we were able to graphically represent the images side-by-side and identify the locations where images have closer keypoints, as shown in Figure 3.

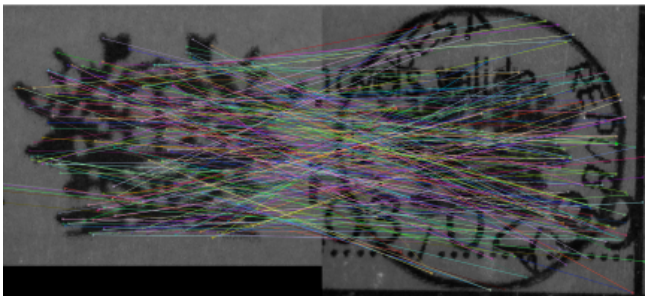


Figure 3: Graphical output representing keypoints matching

It is worth noting that, in Figure 3, both original and stored images of documents do not have perfect cuts around the area of interest to analyse. However, despite these imperfections, the results obtained by the algorithm remained uninfluenced.

With the set of experiments conducted for the Portuguese driver's license cards, we have reached the following results:

Best Candidate	2nd	3rd	4th	5th
83,2%	79,6%	59,2%	48,4%	44,8%

Table 1: Results of experimental setup (similitude degree)

The top 5 best candidates shows that the first 2 have a high degree of similitude, which means that they represented the cases with the most probability of having similar *modus operandi*.

Comparing to the correlation made by the experts of the Forensic Laboratory, the case in study had one other case correlated, and those two images belong to that case.

Therefore, not only the algorithm correctly identified candidates with high values of similitude (over 79%), but also since there was only one case associated, the remaining candidates in the top 5 had low values of similitude (under 60%).

4 Conclusions and Future Work

The purpose of this paper was to introduce a solution able to recognize similar *modus operandi* of frauds in counterfeited documents. To a criminal investigation, this relation can be useful to find the source of the production of the counterfeited documents and to avoid future falsifications of the same type. Nowadays, this recognition is carried out manually, consuming human resources and time.

The presented solution uses visual computing algorithms to compare areas of the documents where a counterfeits are identified.

The current dataset has almost 10.000 images and includes only filters by country and document type. Therefore, the current method is getting more complicated to carry out.

With a filtered dataset, the first version of the presented algorithm took about 70 minutes to process 1500 images. Optimizing the mathematical calculations and introducing parallel computing, the current processing time is about 7 minutes. Given that an expert takes at least 15 minutes to identify the easier falsifications, the current processing time is considered a really good improvement, not only time-wise, but also for relieving human resources. A new parallel computing version is being developed with Graphical Process Unit (GPU), which may improve the time used in each analysis.

Finally, this algorithm was designed to process any kind of official document or image that the forensic laboratories usually work with, e.g., stamps or banknotes.

References

[1] A. Kaur and A. K. G. Vaibhav Saran, "Digital Image Processing for Forensic Analysis of Fabricated Documents", in *International Journal of Advanced Research in Science, Engineering and Techonology*, pp. 84-89, September 2014.

[2] R. Bertrand, P. Gomez-Kramer, O. R. Terrades, P. Franco and J.-M. Ogier, "A System Based On Intrisic Features for Fraudulent Document Detection", in 12th Internation Conference on Document Analysis and Recognition, pp. 106-110, August 2013.

[3] J. Fridrich, D. Soukal and J. Lukás, "Detection of Copy-Move Forgery in Digital Images", in Forensic Science International, vol. 231, pp. 284-295, September 2013.

[4] H. Farid, "Image Forgery Detection", in *IEEE Signal Processing Magazine,* pp. 16-25, March 2009.

[5] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", in 4th Alvey Vision Conference, pp 147-151, 1998.

[6] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", in International Journal of Computer Vision, pp. 91-110, 2004.

[7] H. Bay, A. Ess, T. Tuytelaars and Luc Van Gool, "Speeded-Up Robust Features (SURF)", in Computer Vision and Image Understanding, vol. 110, pp. 346-359, June 2008.

[8] Edward Rosten and Tom Drummond, "Machine Learning for High-Speed Corner Detection", in 9th European Conference on Computer Vision, May 2006.

[9] E. Rublee, V. Rabaud, K. Kunolige and G. Bradski, "ORB: An Efficient Alternative to SIRF or SURF", in IEEE International Conference on Computer Vision, pp. 2564-2571, November 2011.

[10] R. Vieira, C. Silva, M. Antunes, A. Assis, "Information System for Automation of Counterfeited Documents Images Correlation", Procedia Computer Science, 100, 2006, pp. 421-428.

Segmentation of Vascular Networks: A Technological Review

Ricardo J. Araújo[1]
ricardo.j.araujo@inesctec.pt
Jaime S. Cardoso[1,2]
http://www.inescporto.pt/~jsc/
Hélder P. Oliveira[1]
http://www.inescporto.pt/~hfpo/

[1] INESC TEC,
Campus da FEUP, Rua Dr. Roberto Frias, 4200-465,
Porto, Portugal

[2] Faculdade de Engenharia da Universidade do Porto,
Rua Dr. Roberto Frias, s/n, 4200-465,
Porto, Portugal

Abstract

Regarding breast cancer, the mastectomy is still often performed and has even been increasing in some institutions. The removal of the breast(s) is associated with a psychological burden for the patient, but fortunately breast reconstruction is available. The Deep Inferior Epigastric Perforator (DIEP) flap has become the state-of-art for breast reconstruction and it requires preoperative imaging studies. These allow for an expert to extract the characteristics of the different epigastric perforator vessels available in the abdomen of the patient, which are essential for the surgeon to select the most viable one and design the flap. Since the expert analysis is subjective, the preoperative and surgical findings are often different, turning it a nonobjective procedure. Vessel segmentation and characterization algorithms may be able to provide more objective findings. In fact, the literature already contains several methodologies regarding vessel segmentation with other relevant clinical applications. In this paper, we present a brief description of the different methodologies that have been used for 2D/3D vessel segmentation, with the goal of understanding their current limitations.

1 Introduction

Breast cancer is a malignant tumour with origin in the breast tissue, as defined by the American Cancer Society, and it is estimated that more than 230.000 new cases will affect women in the United States during 2016 [7]. The mastectomy is still a highly recurrent procedure and has even been increasing in some institutions [5]. This might suggest that some patients consider the removal of the entire breast a more reliable approach to completely eliminate the tumour. The option to reconstruct the breast afterwards makes this idea more viable.

Reconstruction methods allow to recreate the breast shape, improving the way how women feel about themselves and their image after their breast(s) was(were) removed. The techniques involve the use of implants or tissue from the patient own body. The latter is known as autologous flap and avoids foreign body reactions and leads to more natural and longer lasting results, although it requires more complex and time consuming surgeries [1]. The tissue can be collected from different regions of the body, but the most common site is the belly. The state-of-the-art technique to conduct such reconstruction is the DIEP flap, since it uses microsurgery techniques in order to disturb less the abdominal muscle when compared to the previous state-of-the-art method, the Transverse Rectus Abdominis Muscle (TRAM) flap.

Microsurgery brought the need of performing preoperative imaging studies, where a physician acquires images of the abdominal anatomy of the patient and extracts the characteristics of the available inferior epigastric perforator vessels. Based on a report, the surgeon selects the perforator which seems more adequate for the reconstruction and designs the flap accordingly. The first medical imaging techniques used to study the abdominal perforators were based on Ultrasound probes. Despite their relatively inexpensive cost, they heavily rely on the experience of the physician. Recently, Computed Tomography Angiography (CTA) became the state-of-art method for such task, given the high resolution of the acquired images and physician independence. Nonetheless, efforts are being made to improve the resolution of the images provided by Magnetic Resonance Angiography (MRA) in order to replace CTA, since MRA does not involve radiation [6].

Even though the images acquired with CTA have high resolution, the detection of the inferior epigastric perforators is still a hard task due to their low Signal-to-Noise ratio (SNR). Then, the extraction of measures such as the caliber by the physician have a high subjectivity associated and preoperative findings are commonly different from the surgical ones. This highlights the need of having a method which segments the different perforator vessels and extracts their characteristics in an accurate and objective manner. This clinical application is not the only that can benefit from a Computer Aided Detection system. In fact, other vascular networks have been highly focused by the scientific community in the last years. Retinal 2D images and coronary MRI data are examples of imaging modalities for which several vessel detection and characterization methodologies have been proposed, as they provide important insight about diabetic retinopathy and coronary constriction, respectively, among other pathologies. Hence, this paper presents a brief state-of-the-art review of the main approaches for vessel segmentation, aiming to describe what has been proposed in the literature and the current limitations on the field.

2 State-of-the-art

To the best of our knowledge, there is no algorithm in the literature concerning the segmentation of the inferior abdominal perforators. Still, several authors focused the segmentation of other vascular networks, such as the retinal, coronal and pulmonary vessels. The extraction frameworks are very heterogeneous and it is common to divide them into different categories. In this paper, we present an overview of pattern recognition, matched filtering, vessel tracking, mathematical morphology, multi-scale, region-growing and model-based approaches. For a more thorough description of the existing proposals in each of these categories, consider the reviews of Fraz *et al*. [3] on retinal 2D images and Lesage *et al*. [4] on methods for segmenting 3D vessels.

Pattern recognition techniques perform automatic detection of vascular networks and are divided into two categories, supervised and unsupervised learning methods, depending whether they require prior label information or not. Supervised learning uses the provided *Ground Truth* in order to find a set of rules which are able to distinguish vessels from non-vessel structures, given data extracted features. Although these methods often produce the best results [3], they demand that prior label information is available. As the manual segmentation of data by an expert is a very time-consuming task, it is not a common scenario, especially in real-life applications. On the other side, unsupervised learning is not dependent on this prior information, since it seeks to find inherent patterns of blood vessels on the feature space.

Matched filters are among the first methodologies used to detect vessels (see Figure 1). These methods convolve a previously designed kernel with the data in order to find the desired features. This makes the design of the kernel a crucial step. Regarding this topic, it is common to assume that vessels have limited curvature and may be approximated by piecewise linear segments. Besides, the cross-sectional intensities are usually modeled by a Gaussian distribution, although other distributions have been used, in order to deal with the central reflex problem or more complex cross-sectional profiles. In addition, the cross-sectional intensity of the vessel is dependent of the imaging modality [4]. This family of methods easily presents a high computational cost due to calculating the response over the entire data and at multiple orientations. In addition, there might be the need to consider multiple kernels due to the existence of vessels with different cross-sectional profiles. Handling very tortuous segments is usually not trivial with this kind of approach.

Vessel tracking methods iteratively detect points along the center of a vessel between two given locations, focusing a single segment at each time. New centerline points are estimated by using local information in

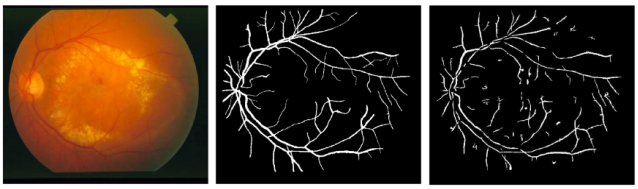


Figure 1: Left to right: pathological retinal fundus image, vessel *Ground Truth* and matched filter based segmentation proposed in [8]. Adapted from [8].

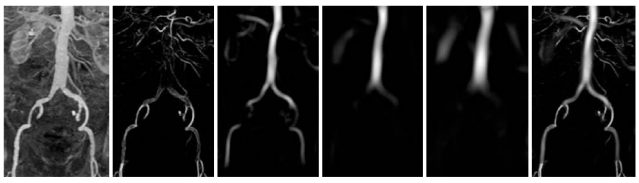


Figure 2: Left to right: aortoiliac MRA, four increasing scale responses and final enhanced image, as proposed in [2]. Adapted from [2].

order to find the path which best matches a certain profile model. This framework is able to naturally present more detailed information on a certain vessel segment than other methods [3]. To detect the vessel boundaries, the pixels/voxels orthogonal to the tracking direction are analyzed. Besides, these methods are usually fast, given that only a portion of the data needs to be analyzed. Nonetheless, they require initialization points which are commonly manually given. Furthermore, not detecting bifurcations will prevent the segmentation of certain portions of the vascular network. Depending on the local features that guide the tracking, recovering from a signal loss is not trivial.

Mathematical morphology theory has also been applied to this topic. It comprises a set of methods that extract objects according to their shape and form. This is accomplished by using structuring elements along with basic operators, such as dilation, erosion, closing and opening. Dilation fills holes and merges disjoint regions, while erosion shrinks objects and eliminates bridges, being the result of both operators controlled by the structuring element. The closing operator is simply a dilation followed by an erosion whilst the opening operator is an erosion followed by a dilation. Although mathematical morphology generally targets binary images, an extension has been made for grayscale data. To enhance vessels, two transforms have been used [3]. The *top-hat* transform consists in subtracting from an image its opened version, being able to enhance tubular structures which are brighter than the background. To enhance darker vessels instead, one should consider the *bottom-hat* transform, which subtracts the image to its closed version. Mathematical morphology based methods are fast and noise resistant but do not consider the cross-sectional information, which can be crucial to distinguish vessels from other structures presenting a local tubular pattern, as sometimes occurs with pathologies.

Multiscale-based approaches are an important class of methods that have been proposed for vessel segmentation, since their formulation naturally accounts for vessels of varying widths. A scale-space representation of the data is made (see Figure 2), in order to explore vessel features at different image resolutions. This allows to implement robust two-stage detection frameworks, where the first stage starts by extracting larger vessels (at lower image resolutions) and the second stage improves the results by extracting smaller vessels (at higher image resolutions) and increases the overall detail of the segmented vascular network. However, it is important to note that some features do not enjoy a well-defined scale-space theoretical framework [4].

Region-growing based methodologies are initialized with a seed point or region that grows by including neighbor pixels that meet a certain criterion. The simplest implementations set a intensity threshold as the inclusion criteria, making them very prone to both false negative (holes inside the object) and false positive (leakage) problems. Even then, region-growing methods present a high efficiency since they employ a sparse search. More complex criteria and strategies using wave propagation techniques to enforce a spatially coherent propagation have been used to improve the baseline results [3].

Deformable models can be separated into parametric and geometric formulations. Regarding the first class, active contour models, also known as *snakes*, are deformable splines that move in the image domain guided by both internal and image forces. Internal forces are related to the prior shape and resist deformation while image forces dictate the features which attract the model. Usually, the edge information is considered. As active contours are heavily dependent of the initialization and parameter tuning, they have a strong application-dependent character [3]. Furthermore, modeling vessel boundaries implies strong elongations of the model, leading to the necessity of creating more complex formulations, such as enabling contour splitting and merging [4]. Geometric deformable models are based on the theory of curve evolution geometric flows and are frequently implemented using the level-set method. Its advantage relies on avoiding the parameterization of the objects by performing computations in a fixed Cartesian grid. However, computational cost is increased and special algorithmic care has to be taken to ensure convergence [4].

3 Conclusion

Vessel segmentation has been a widely focused topic in the literature in the last years, as computer vision methods are able to support the activity of physicians who analyze vascular networks in their daily lives. Although different methodologies have been proposed to address such task, they are commonly just adequate for a specific vasculature and/or imaging modality. Besides, only a portion of the algorithms can cope with problems such as central reflex and the detection of low SNR vessel segments remains a major challenge. Thus, the development of algorithms that are able to accurately detect general vascular networks, including low SNR vessels, are highly welcome.

Acknowledgements

This work was funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] Breastcancer.org. `www.breastcancer.org`. Accessed: 2016-07-07.

[2] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Proc. Medical Image Computing and Computer-Assisted Intervention*, pages 130–137, 1998.

[3] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. Blood vessel segmentation methodologies in retinal images - a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407–433, 2012.

[4] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea. A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. *Medical Image Analysis*, 13(6):819–845, 2009.

[5] U. Mahmood, A. L. Hanlon, M. Koshy, R. Buras, S. Chumsri, K. H. Tkaczuk, S. B. Cheston, W. F. Regine, and S. J. Feigenberg. Increasing national mastectomy rates for the treatment of early stage breast cancer. *Annals of Surgical Oncology*, 20(5):1436–1443, 2013.

[6] G. F. Pratt, W. M. Rozen, D. Chubb, M. W. Ashton, A. Alonso-Burgos, and I. S. Whitaker. Preoperative imaging for perforator flaps in reconstructive surgery. *Annals of Plastic Surgery*, 69(1):3–9, 2012.

[7] R. Siegel, K. Miller, and A. Jemal. Global cancer statistics. *A Cancer Journal for Clinicians*, 65(1):5–29, 2015.

[8] B. Zhang, L. Zhang, L. Zhang, and F. Karray. Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in Biology and Medicine*, 40(4):438–445, 2010.

Vessel width estimation in eye fundus images

Teresa Araújo[1,2]
bio11052@fe.up.pt
Ana Maria Mendonça[1,2]
amendon@fe.up.pt
Aurélio Campilho[1,2]
campilho@fe.up.pt

[1] Faculdade de Engenharia da Universidade do Porto
Porto, Portugal
[2] INESC-TEC-INESC Tecnologia e Ciência
Porto, Portugal

Abstract

Changes in the retinal vessel caliber are associated with several major diseases and can be evaluated using eye fundus images. However, the clinical assessment is tiresome and prone to errors, motivating the development of automatic methods.

A method based on vessel intensity profile model fitting for the estimation of retinal vessel widths is proposed in this work. A new parametric model is used, consisting in a Difference-of-Gaussians multiplied by a line which is able to describe profile asymmetry. The parameters of the best-fit models are used to determine the vessel widths using ensembles of bagged regression trees.

The method is evaluated on the public dataset REVIEW, showing a precision close to the observers, outperforming other state-of-the-art methods. The method is robust and reliable for width estimation in images with pathologies and artifacts.

1 Introduction

Retinal vessels are the only portion of the circulation that is directly observable [1]. Changes in retinal vessel caliber, which can be evaluated using high resolution eye fundus color images, are an important sign of major diseases, such as diabetes, hypertension, arteriosclerosis, cardiovascular diseases, pre-diabetes and pre-hypertension [6, 10]. The early diagnosis is crucial to prevent and reduce health damages. Automated measurement methods are desirable, in order to improve the efficiency and reliability of the results, namely in screening programs, considering the large number of images and the complexity of the vascular network. The development of automatic methods for width measurement is challenging, due to the variability of the vessels appearance, image quality and resolution as well as the lack of standardized data for algorithms comparison [8]. Most of the state-of-the-art methods present limitations, such as poor performance in lower resolution images, on thinner vessels or susceptibility to artifacts and pathologies. The objective of this work is the development of an automatic method for robust estimation of vessel caliber in eye fundus images, improving the current state-of-the-art results, particularly in the most difficult cases of small blood vessels and images with pathologies and artifacts. The method is evaluated on a public dataset designed for vessel width measurement evaluation and compared with the state-of-the-art methods using adequate performance metrics.

2 Materials and Methods

2.1 Vessel segmentation

The first step of the algorithm is the vasculature segmentation. In this work we apply the method proposed by [9]. This method combines centerline detection and region growing for the segmentation of retinal blood vessels. Fig. 1-b) shows an example of the application of the segmentation algorithm to a region of an eye fundus image from the REVIEW dataset.

2.2 Vessel centerline detection and segment extraction

Vessel centerlines are detected from the binary vessel image. For this purpose, the skeletonization of the image is performed through thinning [7] (Fig. 1-c)). Then, bifurcation and crossover points are excluded by removing pixels with three or more neighbors (8-connection scheme). Segments are refined by removing short spurs that may have resulted from the thinning process [3].

2.3 Vessel intensity profile extraction

The extraction of the cross-sectional intensity profiles requires the determination of the vessel orientation. Spline approximation is applied to smooth the segments [3]. The first derivatives of the splines are computed and the vessel normals determined based on the vessel directions. Intensity profiles are extracted along the normals, on the green channel of the image, due to its larger contrast. Fig. 2-b) shows a surface constituted by smoothed 1D profiles extracted from the segment of Fig. 2-a), stacked together in parallel to each other, aligned by their centers. The region of the profile containing the vessel in study is identified using a method based on peak search [8]. Then, profiles are smoothed using Anisotropic Gaussian filtering, which allows the application of more smoothing in the direction of the vessel than perpendicularly, avoiding excessive blurring of the vessel edges [3].

2.4 Model fitting

The intensity profiles are then approximated by a model through finding the parameters that lead to the best adjustment between the curve and the profile. Two models are tested: Hermite model with 6 parameters [8] and a newly proposed DoG-based model. Both models consider the existence of central light reflex (CLR), which consists in an elevation of the intensity profile in its center region. 2D model fitting is applied to the set of points retrieved from 11 neighboring vessel cross-sectional intensity profiles, since the 2D approach is more robust than the 1D one, introducing some smoothness in the fitting. The new model consists in a Difference-of-Gaussians multiplied by an inclined line which can module the vessel profile asymmetry. This model is given by:

$$m(x,y) = (t + h_1 \times e^{-(\frac{x-\mu}{\sqrt{2}\times\sigma_1})^2} - h_2 \times e^{-(\frac{x-\mu}{\sqrt{2}\times\sigma_2})^2}) \times (\lambda \times (x-\mu) + t) \quad (1)$$

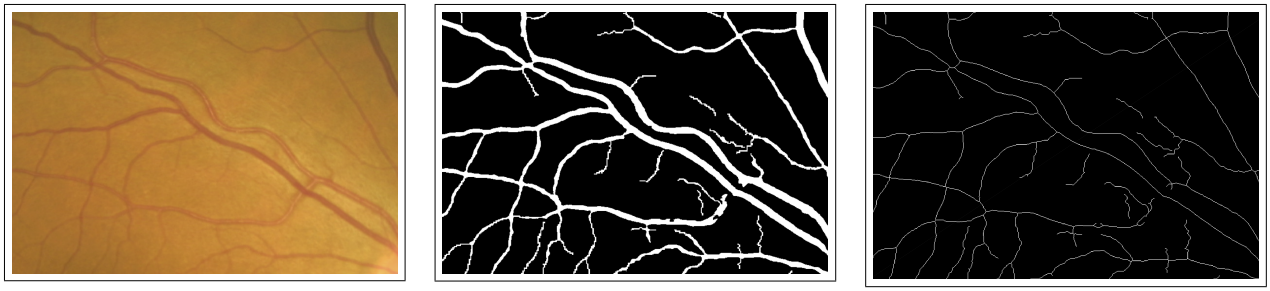
where x is the coordinate along the vessel cross-section, t is the maximum of the function, h_1 is the height of the first (main) Gaussian, μ the location of the center, σ_1 the spread of the first Gaussian, h_2 the height of the second Gaussian, σ_2 the spread of the second (CLR) Gaussian and λ is the slope of the multiplying line. The means of the two Gaussians are the same, positioning the light reflex in the center of vessel. The parameters of the best-fit-model are found by solving a non-linear least squares problem, using the Trust-Region-Reflective method [5]. The allowed range of parameters and the parameter initialization are previously defined based on the common appearance of the vessel profiles. An example of fitting to a vessel profile using the proposed model is shown in Fig. 2-c).

2.5 With estimation

Once the best-fit model is found, the relationship between its parameters and the vessel width is determined. In this work, this relation is obtained using ensembles of bagged decision trees [8]. Ensemble methods, such as bagging, i.e., bootstrap aggregation, combine multiple weak trees, forming a more accurate and robust regressor [4]. Additionally, we use random forests, being that each tree in the ensemble can randomly select predictors for the decision splits, improving the accuracy of the predictions.

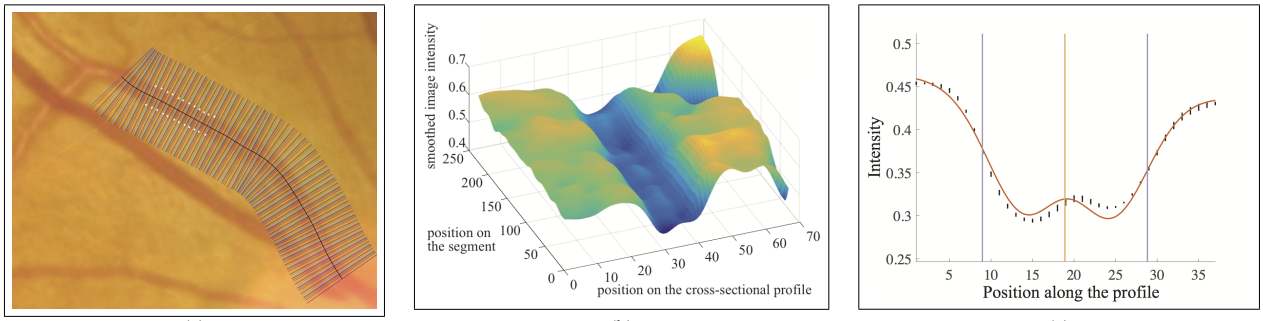
3 Results

The REVIEW (Retinal Vessel Image set for Estimation of Widths) dataset [2] is the only public dataset with vessel width measurements, marked by 3 independent observers on randomly selected segments. It has 4 sets, 16 images, 193 segments and 5066 profiles. These images have a variety of



(a) (b) (c)

Figure 1: Vessel segmentation and centerline detection: (a) CLRIS001 region; (b) Segmented vessels; (c) Vessel centerlines.



(a) (b) (c)

Figure 2: Vessel intensity profile extraction, smoothing and model fitting: (a) Normals to the vessel; (b) Smoothed extracted intensity profiles; (c) Profile model fitting. Dots in a) correspond to the ground truth. Vertical ines in c) represent the vessel centerline and limits accordingly to the ground truth annotations.

Table 1: Performance of vessel width measurement methods on the REVIEW dataset, for each of the 4 sub-datasets (HRIS, VDIS, CLRIS and KPIS), in terms of standard deviation of the errors (pixels). HRIS: The high resolution image set; VDIS: The vascular disease image set; CLRIS: The central light reflex image set; KPIS: The kick point image set.

Method	HRIS	VDIS	CLRIS	KPIS
Observer 1 | 2 | 3	0,288 0,256 0,285	0,543 0,621 0,669	0,567 0,698 0,566	0,233 0,213 0,234
Lupascu [8]	0,438	1,073	1,154	0,318
Bankhead [3]	0,32	0,95	0,95	0,29
Proposed (Hermite)	0,221	0,726	1,152	0,283
Proposed (DoG-L7)	0,217	0,690	0,563	0,298

resolutions, pathologies and artifacts. The ground truth is the mean of the annotations of the 3 observers [2]. In this context, it is more relevant that the algorithms retrieve precise results, i.e., low standard deviation of the width errors, than accurate, i.e., low mean of the errors [2]. This is true since any consistent bias (nonzero mean of the errors) can be compensated by the subtraction/division of a bias, whereas fluctuations of the errors (nonzero standard deviation) cannot.

Table 1 shows the results of our method and of relevant and recent state-of-art methods, in terms of standard deviation of the width errors. Our method is evaluated using 10-fold cross-validation inside each dataset [8]. Considering the proposed method, the best results are obtained with the introduced model, specially for CLRIS dataset, for which it shows approximately twice the performance of the Hermite model. Comparing to the other methods, ours has consistently the highest precision. Our algorithm using the proposed model practically halves the error reported by [8]. The obtained results are good due to the use of an appropriate parametric model but also due to the application of several preprocessing steps before model fitting which improves the vessel width measurement performance.

4 Conclusions

Our retinal vessel width measurement method often outperforms other state-of-the-art methods, retrieving precise results, close to that of the observers, as was the goal. This shows the robustness of the method and its great potential as a tool for automatic measurement of retinal vessel widths. Despite this, there is still room for improvement and adaptations. The definition of a better training dataset, with a more uniform width distribution, is expected to increase the reliability of the results. Further optimization of the algorithm can be performed, by improving steps such as the initial profile length estimation and the ensemble parameters.

Acknowledgements

Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016" is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] M. D. Abramoff, M. K. Garvin, and M. Sonka. Retinal Imaging and Image Analysis. *IEEE Transactions on Medical Imaging*, 3:169–208, 2010.

[2] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry. REVIEW - a reference data set for retinal vessel profiles. In *Conference proceedings : 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*, pages 2262–2265, 2008.

[3] P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLoS ONE*, 7(3):1–25, 2012.

[4] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[5] T. Coleman, M. A. Branch, and A. Grace. *Optimization Toolbox For Use with M ATLAB*. Matlab The Mathworks Inc, 1999.

[6] M. K. Ikram, Y. T. Ong, C. Y. Cheung, and T. Y. Wong. Retinal Vascular Caliber Measurements: Clinical Significance, Current Knowledge and Future Perspectives. *Ophthalmologica*, 229(3):125–136, 2013.

[7] L. Lam, S.-W. Lee, and C.Y. Suen. Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (9):869–885, 1992.

[8] C. A. Lupascu, D. Tegolo, and E. Trucco. Accurate estimation of retinal vessel width using bagged decision trees and an extended multiresolution Hermite model. *Medical Image Analysis*, 17(8):1164–1180, 2013.

[9] A. M. Mendonça, B. Dashtbozorg, and A. Campilho. Segmentation of the Vascular Network of the Retina. In *Image Analysis and Modeling in Ophthalmology*, pages 85–110. 2014.

[10] T. T. Nguyen, J. J. Wang, and T. Y. Wong. Retinal Vascular Changes in Pre-Diabetes and Prehypertension. *Diabetes Care*, 30(10):2708–2715, 2007.

Human Pose Estimation Using Wide Stacked Hourglass Networks

Miguel Farrajota
mafarrajota@ualg.pt
J.M.F. Rodrigues
jrodrig@ualg.pt
J.M.H. du Buf
dubuf@ualg.pt

Vision Laboratory, ISR (Lisbon), LARSyS,
University of the Algarve
Campus de Gambelas, 8005-139 Faro, Portugal

Abstract

Pose estimation is the task of predicting the pose of an object in an image or in a sequence of images. Here, we focus on articulated human pose estimation in scenes with a single person, by extending the stacked hourglass convolutional network architecture. In this network topology, features are processed across all scales capturing the various spatial relationships associated with the body, by employing repeated bottom-up and top-down processing (used with intermediate supervision which is a key step in this architecture). We propose some improvements to further increase performance, namely: (a) increase the depth and the number of bottom-up, top-down processing stacks, with (b) increasingly wider residual blocks to increase the networks prediction accuracy. We demonstrate top-performing results on the popular FLIC dataset.

1 Introduction

Human pose estimation has substantially progressed recently on many popular benchmarks [1, 4, 6], including single person pose estimation [2, 8, 9, 10, 11, 13]. For a pose estimation system to be effective it must be robust to occlusion, deformation, sufficiently accurate on rare and novel poses, and invariant to changes in appearance due to factors like clothing and lighting. Early work on pose estimation tackled these difficulties by using robust image features and sophisticated structured prediction [10]. Deep learning methods [9] have replaced the conventional pipeline by the use of convolutional neural networks (ConvNets) which constitute the main driver behind the huge rise in performance of many computer vision tasks. Recent pose estimation systems [2, 8, 11, 13] have adopted ConvNets as their main building block, completely replacing hand-crafted features and graphical models.

In this paper we extend the work of Newell et al. [8] by using wider residual blocks with more feature planes and employing more and deeper hourglass stacks to the network. This type of network allows to capture information across all scales of the image. It is composed of consecutive steps of pooling and up-sampling to get the final output of the network. The hourglass network pools down to a very low resolution and then up-samples and combines features from multiple resolutions. This topology allows for repeated bottom-up, top-down inference across scales, which in conjunction with the use of intermediate supervision, is critical to the final performance. By using more residual blocks per network, with increasingly wider convolutional feature planes as the number of hourglass stacks increases, we improve the original [8] network's base prediction capability to predict body joint positions.

The main contribution of this paper is to provide an insight of the use of deeper and wider residual blocks with more top-down, bottom-up processing stacks with intermediate supervision in order to increase performance when compared with the original stacked hourglass network configuration in [8]. The resulting analysis shows improvements in accuracy for the FLIC dataset [4], with 99.1% for wrists and 97.9% for elbows.

2 Pose estimation

The human poses estimation scheme works as follows: (i) the model takes as input an image with a centered person and outputs a heatmap of all body joints; then (ii) the final prediction of the network consists of extracting the max activating location of the heatmap for any given joint.

The present scheme uses the same architecture as in [8] with some core modifications. Our architecture is composed of four connected hourglass networks that each output a heatmap of the body joints into the next one. The hourglass networks are composed of consecutive convolutional and max pooling layers that are used to process features down to a very low resolution. At each max pooling step, the network branches off and applies more convolutions at the original unpooled resolution. After reaching the lowest resolution, the network then up-samples and combines features across scales. To connect information from two adjacent resolutions, we do nearest neighbor up-sampling of the lower resolution followed by an element-wise addition of the two sets of features. Here, the hourglass topology is symmetric, meaning that for every layer on the way down (reducing resolution) there is a corresponding layer going up (increasing resolution). All convolutions used in the hourglass network consist of bottlenecked residual blocks composed of one 1×1 convolutions to reduce dimensionality followed by one 3×3 convolutions and by another 1×1 convolutions to increase dimensionality. Also, as the number of hourglass network stacks increases, we increase the network's convolutions feature planes such that the last hourglass networks have more parameters than the previous ones on the stack. After reaching the output resolution of the network, two consecutive rounds of 1×1 convolutions are applied to produce the final network predictions. The output of the network consists of a set of heatmaps for each body joint where the network predicts the probability of a joint's presence for each pixel. The final prediction of the network is the max activating location of the heatmap for a given joint. Figure 1 (top) shows the network's architecture.

The proposed method is trained and evaluated on the FLIC dataset [4]. This is a relatively small and popular benchmark. It contains 5003 images from popular Hollywood movies with annotated upper body parts and with most persons facing towards the camera. We used a standard train/test split, with 80% of the images being used for train, and the remaining 20% for test. We followed the evaluation protocol proposed by [10] and evaluated the wrist and elbow body parts using the Percentage of Correct Keypoints (PCK) metric, which reports the percentage of detections that fall within a normalized distance of the ground truth keypoints.

The network was trained using Torch7 [3] and optimized using rmsprop with a learning rate of 0.0002.5, alpha of 0.99 and epsilon of $1e-8$. All network weights were randomly initialized with a uniform distribution. We used mini-batches of 4 randomly sampled person poses centered in the input image. This is achieved by centering along the x-axis by using the torso bounding box annotation. Input images were resized and cropped to 256×256 pixels. We performed data augmentation by applying rotation to the sample images ($\pm 30^{\circ}$) and scaling (.75-1.25). Batch normalization [5] was used for faster convergence during training.

During back-propagation we used the same technique as [8, 11], where a Mean-Squared Error (MSE) is applied for comparing the predicted heatmap to a ground-truth heatmap, consisting of a 2D gaussian centered on the joint location with standard deviation of 1 pixel. To improve performance at high precision thresholds, the prediction is slightly offset by a quarter of a pixel in the direction of its next highest neighbor before transforming back to the original coordinate space of the image.

3 Discussion and results

We presented a method for human pose estimation with a wide stacked hourglass network (Fig. 1, top row). The proposed method employs a modified hourglass network with more stacks and wider feature maps than [8], resulting in better detection performance. Results showed that deeper networks containing more convolutions with more feature channels combined in a network with several stacked hourglass networks results in a noticeable increase in accuracy, but also in inference time.

We show results of the method's performance on detecting poses of individual persons. Figure 1 (middle row) shows heatmaps of body joints

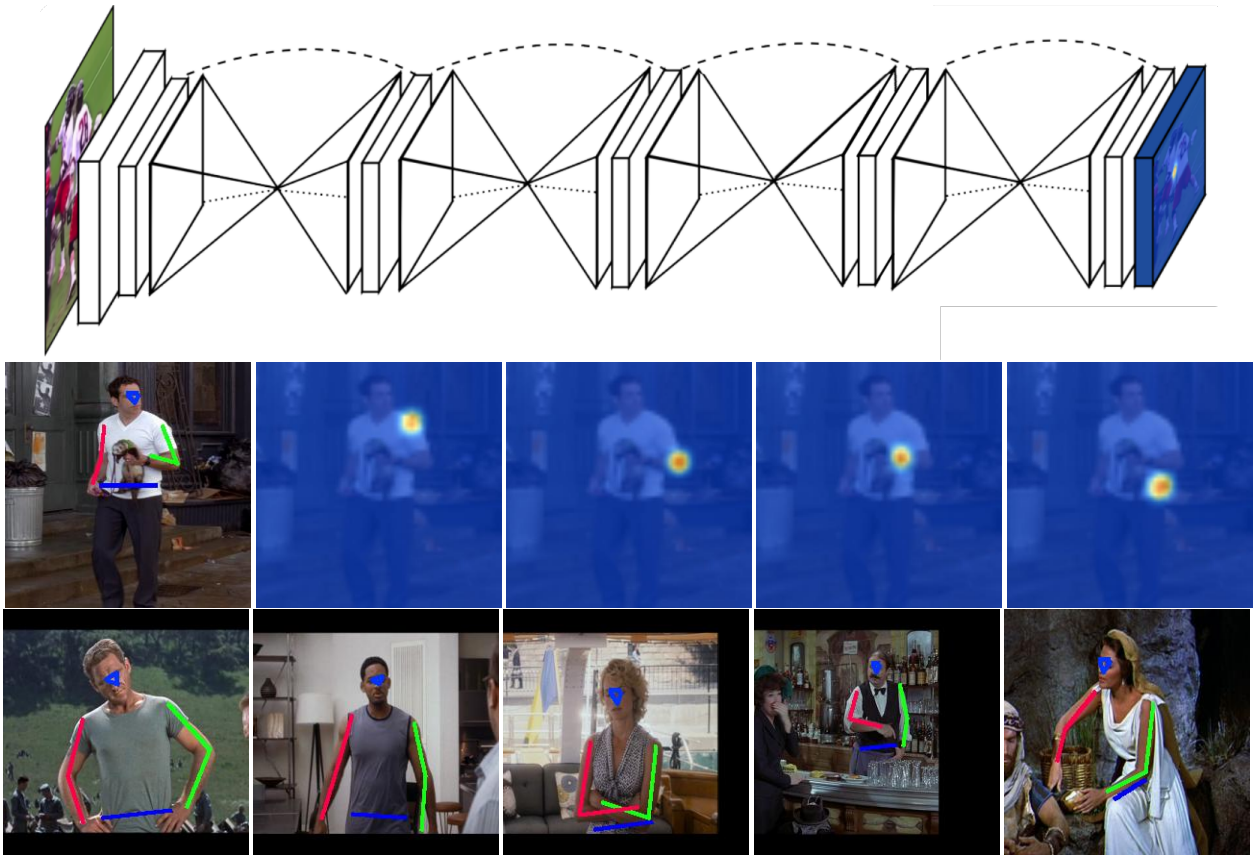


Figure 1: Network's architecture illustration (top, adapted from Fig.1 in [8]) with detection heatmaps (middle) and detection results on the FLIC dataset [4] (bottom).

Methods	Elbow	Wrist
Sapp et al. [10]	76.5	59.1
Toshev et al. [12]	92.3	82.0
Tompson et al. [11]	93.1	89.0
Chen et al. [2]	95.3	92.4
Wei et al. [13]	97.6	95.0
Newel et al. [8]	99.0	97.0
Ours	**99.1**	**97.9**

Table 1: Performance comparison our method and other 6 methods on the FLIC dataset evaluation protocol (PCK@0.2) [10].

of the left side of a person, and (bottom row) some detection results on test images of the FLIC dataset [4]. Table 1 shows some benchmark results on the FLIC dataset of our network and other top-performing methods. We report accuracy using the metric introduced in Sapp et al. [4] for the elbow and wrist joints. Our method shows competitive results, reaching 99.1% PCK@0.2 accuracy on the elbow and 97.9% on the wrist.

In future work we expect to benchmark our method with more popular datasets like MPII [1] and Leeds Sports [6] for single person detection and MSCOCO 2016 keypoint challenge [7] for multiple persons in a scene.

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] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2D human pose estimation: New benchmark and state of the art analysis. In *CVPR*, pages 3686–3693, 2014.

[2] Xianjie Chen and Alan L Yuille. Articulated pose estimation by a graphical model with image dependent pairwise relations. In *NIPS*, pages 1736–1744, 2014.

[3] Ronan Collobert, Koray Kavukcuoglu, and Clément Farabet. Torch7: A Matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, number EPFL-CONF-192376, 2011.

[4] Andreas Ess, Bastian Leibe, Konrad Schindler, and Luc Van Gool. A mobile vision system for robust multi-person tracking. In *CVPR*, pages 1–8, 2008.

[5] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[6] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *BMVC*, volume 2, page 5, 2010.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, pages 740–755. Springer, 2014.

[8] Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked hourglass networks for human pose estimation. *arXiv preprint arXiv:1603.06937*, 2016.

[9] Leonid Pishchulin, Mykhaylo Andriluka, Peter Gehler, and Bernt Schiele. Strong appearance and expressive spatial models for human pose estimation. In *ICCV*, pages 3487–3494, 2013.

[10] Ben Sapp and Ben Taskar. Modec: Multimodal decomposable models for human pose estimation. In *CVPR*, volume 13, page 3, 2013.

[11] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *CVPR*, pages 648–656, 2015.

[12] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *CVPR*, pages 1653–1660, 2014.

[13] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. *arXiv preprint arXiv:1602.00134*, 2016.

Semantic Modelling for User Interaction with Sonic Content

António Sá Pinto*
antoniosapinto@gmail.com
Matthew E.P. Davies
mdavies@inesctec.pt
Perfecto Herrera
perfecto.herrera@upf.edu

Faculdade de Engenharia da Universidade do Porto
Porto
Sound and Music Computing Group, INESC TEC
Porto
Music Technology Group, Universitat Pompeu Fabra
Barcelona

Abstract

In this paper we present a methodology for converting semantic descriptions of sounds into computable audio features. This process aims to enable the use of commonly used notions of timbre in an audio engineering context where the user interacts (e.g. searches for sounds in large digital collections) with sonic content, bridging the gap between the high-level perceptual sound notions and low-level machine-ready descriptors. Although our focus is the description of the constituent blocks of a general-purpose semantic framework, examples from an experimental test for the semantic characterization of drum samples will be given to illustrate the process.

1 Introduction

The semantic gap that separates human descriptions of sounds from computable definitions is a recognized topic in the Music Information Retrieval (MIR) field of research [9]. Due to escalating data availability, and the potential and natural appeal of using common vocabulary for interacting with this abundance of sound resources, semantic audio studies have been increasingly addressed by the MIR scientific community. Accordingly, multiple applications have been proposed, such as active listening [11], music recommendation [2], sound retrieval [12] or intelligent audio production [10].

Our focus is the use of semantic descriptors (adjectives) for user interaction with sonic content, in a music production environment, where tasks such as browsing, retrieval or identification of samples are typical. Rather than detailing a specific system or application, we aim at identifying the building blocks required to enable the use of (high-level) human vocabulary by the user, dismissing the need for the extensive training required for interaction with sonic content in the (low-level) machine-extractable acoustic space.

2 Methodology

In this section we explain the proposed method for mapping semantic notions of sound into acoustic-based computable parameters, and the underlying procedures of extracting the audio signals features and creation of a semantic lexicon (a three-tier scheme is presented in Figure 1). Our focus lies on the description of a general methodology, but in this paper we illustrate the approach via the semantic characterization of drum samples.

2.1 The Acoustic (and Psycho-Acoustical) Space

Audio signals are computationally described in terms of audio features. In the case of isolated notes of musical instruments, for each of a set of samples (A_1,...,A_n), a vector of attributes (F_1,...,F_n) is obtained, which provides a compact representation of each of the sonic instances. Audio feature extraction is a cornerstone of audio signal processing, thus a consolidated body of work has been produced in this area (e.g. [8]). This process is achieved through the application of digital signal processing techniques directly to the raw audio signal, whether in the time-domain, wavelet, constant-Q or other spectral domains. There are countless audio features, several domain-oriented taxonomies, which group descriptors by their nature under distinct categories or related standpoints: spectral vs temporal, attack vs sustained vs decay, energy-related, perceptual, etc. Feature Selection and Transformation are normally the following processing blocks, aiming to refine and adapt the feature extraction process to the subsequent tasks in a specific target-application.

*The reported work was part of the Master's thesis done in the MTG-UPF.

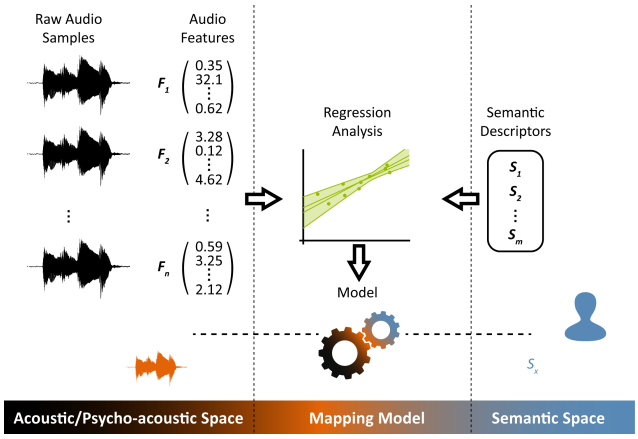


Figure 1: General Methodology Scheme

In our study, we use the 48 features which are fitted to percussive instruments [5]. In addition, four global descriptors were added to our final set: the average and variance of the Spectral Centroid and Spectral Flux, due to their reported relevance in the perceptual (and consequently semantic) space [7].

2.2 The Semantic Space

Terms such as *bright* or *warm* are common to experts when describing the sound of an instrument, and are among reported collections of verbalizations for the description of timbre of musical instruments [4]. However in common usage, such terms are less precise in their meaning, and while this may not prejudice the comprehensibility of a conversation about timbre, it disrupts the development of a computational semantic-based system. Therefore, to accomplish this goal, a consistent lexicon of semantic descriptors must be developed, in order to establish what will become the interface "language" between the user and the system: expert verbalizations of sound descriptions must be collected, and carefully reduced to a consistent subset, whose reliability must be confirmed through appropriate statistical validation techniques (e.g. the Chronbach's Alpha Coefficient).

Despite the existence of several instrument-specific studies, a significant gap still exists in the literature for percussive instruments. Only two systematic studies addressed the constitution of a drum timbre lexicon have been made [1, 3]; nevertheless, they did not establish a body of percussive semantic adjectives commonly accepted by the community as a reference lexicon. Given this absence, an initial collection of timbral adjectives was built upon the referred percussive lexicons, complemented by terms collected from other sources (e.g. e-commerce drum samples websites and an inquiry addressed to percussionists). From this large collection (more than 700 terms), a final subset of five sonic attributes was obtained as a "common denominator" from these several sources (preceding the reliability measurement, the adopted criteria was the unambiguity of their meanings and the coverage of salient sonic perception aspects, balanced with the additional statistical processing effort): (1) Brightness: the quality in sound of being clear, vibrant, and typically high-pitched; (2) Hardness: the quality in sound of being firm, rigid, stiff; (3) Tone (Sensation): the sound provokes a tonal sensation (pitch); (4) Size: the apparent external size, form of the sound source; (5) Ambiance: the conditions or atmosphere in which the sound was produced are explicit (e.g. reverb).

2.3 Mapping the Semantic and Acoustic Space

The final component for a semantic-based system is the one that maps the two parameter spaces: the semantic space, which reflects the user-perceived prominent timbral aspects, and the underlying acoustic and psycho-acoustic features extracted from the sound samples. Given this pivotal role, global design decisions are hereby assembled and disclosed by a thorough analysis: e.g. the suitability of the set of features chosen to characterize the acoustic space or the reliability of the group of semantic descriptors. From its examination, we may draw key findings, namely which sonic cues play a relevant role for each semantic descriptor.

At this stage, we aim to model the relationship between a dependent variable (each of the semantic descriptors) and a group of independent variables (the audio descriptors):

$$S = f(A) \tag{1}$$

For this purpose, a regression analysis has been applied. These techniques provide a set of coefficients for a function that best fits predefined data observations, thus mapping acoustic features into the semantic space. Formally, given $(x_i, y_i), i \in 1,...,N$ a set of N pairs, where x_i is a $1 \times M$ feature vector and y_i is the real semantic rate value to predict, a regressor r is defined as the function that minimize the mean squared error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - r(x_i))^2 \tag{2}$$

We tested four different state-of-the-art regression techniques: Linear, Support Vector, Random Forest and Radial Basis Function (RBF). A brief performance analysis will be presented in the following sub-section. As an example of the mapping model (eq. 1), we highlight the five most significant terms of the Linear Regressor for the Brightness descriptor.

$$\begin{aligned} Brightness = & 3.6 * dSpecCentroid + 2.9 * dSkewness + \\ & 2.3 * dKurtosis - 1.8 * mfcc_{ave05} - 1.8 * RED5 \end{aligned} \tag{3}$$

where $dSpecCentroid$ represents the audio descriptor decay spectral centroid, $dSkewness$ the decay skewness, $dKurtosis$ the decay kurtosis, $mfcc_{ave05}$ the average of the 5th band (of 13) of the mel-frequency cepstrum coefficients (MFCC) representation of the signal spectral envelope, and $RED5$ represents the 5th (of 8) band energy relative percent of the signal, as defined in [5].

2.4 Experimental Results

In order to validate the proposed methodology, a semantic listening experiment applied to drum samples was undertaken, by means of a verbal attribute magnitude estimation (VAME) questionnaire [6], in which 47 (expert) subjects were asked to rate a set of 30 percussive samples (A_1,...,A_{50}), using a 6-point ordinal scale for each of the 5 semantic verbalizations (S_1,...,S_5). In parallel, the aforementioned set of acoustic descriptors (F_1,...,F_{52}) were computed for each audio signal, and after feature transformation and selection, different sets were included for evaluation. Following a 10-fold cross-validation procedure, the regressor's performance evaluation was obtained in terms of R^2 index (the squared correlation coefficient, a standard metric for measuring the accuracy of the fitting of the regression models) and the correspondent MSE. The accuracy was promising in many cases (reaching values of 0.695 for the R^2 score); yet, distinct improvement paths pave the way for a robust model to be built (e.g. coping with the significant inter-rater disagreement). After an informal comparison to other related studies [13], an auspicious overall performance was confirmed.

Although a general performance trend could be identified (a slight pre-eminence of the Random Forest, followed by the RBF), it is not yet possible to define a regression method that well suits all the high-level descriptors, given its dependency on several concurrent issues, such as the audio features set or the selected semantic descriptors. This initial analysis highlights several important issues for consideration in future work, namely the cross-dependency on other experiment design-related factors: the unambiguity of the adjectives (e.g. Hardness and Tone codings were inconsistent among raters), as well as the quality (or expertise) of the ratings (a selection of the top-5 raters, selected under a thorough correlation analysis, approximately doubled one regressor's accuracy). On the other hand, our results infer the suitability of the set of features used to characterize the sound samples. Addressing a main goal of our experimental study (mapping the acoustic and the semantic space), our framework provided some expected results in line with the literature: the relevance of the spectral centroid to Brightness perception, the association between the descriptor Tone and MFCC representation, and the relevance of attack energy and log-attack time to the Hardness descriptor.

3 Conclusions

In this paper we have presented in the context of a practical application, a methodology for converting expert sound notions into computable definitions. This enables a semantic approach for the description of sonic content, that tries to approximate human perception when describing timbre using common adjectives. A general framework architecture has been proposed, and their processing blocks have been characterized. Some experimental examples were given to illustrate the methodology, and aided the discussion of some prominent issues; related both to procedure and implementation viewpoints. We were able to infer some relationships between semantic and acoustic descriptors, confirming some findings reported in reference timbre studies, while validating the presented methodology. Over our study span (here resumed), some interesting paths to further research have been identified (e.g. the use of onomatopoeias to describe percussive instruments, or the application of deep learning methods for our framework). In conclusion, the "semantic gap" represents an important obstacle to overcome, thereupon we foresee a key role for semantic approaches that target a wide range of relevant applications for interaction with sonic content.

References

[1] R. Bell. *PAL : The Percussive Audio Lexicon*. Doctoral dissertation, Swinburne University of Technology, Melbourne, Australia, 2015.

[2] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambó, E. Gómez, and P. Herrera. Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing and Management*, 49(1):13–33, 2013.

[3] W. Brent. *Physical and Perceptual Aspects of Percussive Timbre*. Doctoral dissertation, University of California, San Diego, 2009.

[4] A. C. Disley, D. M. Howard, and A. D. Hunt. Timbral description of musical instruments. In *International Conference on Music Perception and Cognition*, pages 61–68, 2006.

[5] P. Herrera, J. P. Bello, G. Widmer, M. Sandler, O. Celma, F. Vignoli, E. Pampalk, P. Cano, S. Pauws, and X. Serra. SIMAC: Semantic Interaction with Musical Audio Content. In *The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology (EWIMT)*, pages 399–406, 2005.

[6] R. Kendall and E. Carterette. Verbal attributes of simultaneous wind instrument timbres: I. von bismarck's adjectives. *Music Perception*, 10(4):445–467, 1993.

[7] S. Lakatos. A common perceptual space for harmonic and percussive timbres. *Perception & psychophysics*, 62(7):1426–1439, 2000.

[8] G. Peeters. A large set of Audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, 2004.

[9] X. Serra, M. Leman, and G. Widmer. A Roadmap for Sound and Music Computing. Technical report, 2007.

[10] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. Reiss. SAFE: A System for the Extraction and Retrieval of Semantic Audio Descriptors. In *Late Breaking Demo Session, 15th ISMIR Conference*, 2014.

[11] F. Thalmann, A. P. Carillo, G. Fazekas, G. A. Wiggins, and M. B. Sandler. The Semantic Music Player: A Smart Mobile Player Based on Ontological Structures and Analytical Feature Metadata. *Proceedings of the 2nd Web Audio Conference (WAC-2016)*, 2016.

[12] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Trans. Audio, Speech, Language Process.*, 16(2):467–476, 2008.

[13] M. Zanoni, F. Setragno, F. Antonnaci, A. Sarti, G. Fazekas, and M. Sandler. Training-based Semantic Descriptors modeling for violin quality sound characterization. In *Audio Engineering Society Convention 138*. Audio Engineering Society, 2015.

Twitter classification: are some examples better than others?

Joana Costa[1,2]
joana.costa@ipleria.pt,joanamc@dei.uc.pt
Catarina Silva[1,2]
catarina@ipleiria.pt,catarina@dei.uc.pt
Mário Antunes[1,3]
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

One of the major challenges in dynamic environments is the amount of data, specially when dealing with streams. It is sometimes unfeasable to store all the previously seen data, despite the fact that it may carry substantial information for future use. Two questions arise: (i) How is it possible to enhance the input examples? (ii) Are there examples better than others, that thus should be kept for future use?

In this paper we propose a method that determines the most relevant examples by analysing their behaviour when defining separating planes between classes. We have tested our approach in a Twitter scenario and results show that keeping those examples improves the classification performance.

1 Introduction

Social networks have settled definitely in the daily routine of Internet users. They have also gained increasing importance and are being widely studied in many fields of research over the last years, such as computer, social, political, business and economical sciences. With millions of daily users, they are an important source of information and learning in those environments can have multiple benefits, like market sensing, recommendation, event detection, sentiment analysis, among others.

Considering their potential in information spread, it is imperative to find learning strategies able to learn in social networks. However, their dynamic nature, requires specific learning approaches. Differently from the commonly used approaches, effective learning in such scenarios requires a learning algorithm with the ability to detect context changes without being explicitly informed about them, quickly recovering from those context changes and adjusting hypothesis to new contexts. Multiple drift patterns were identified by Zliobaite [1], namely sudden, gradual, incremental, and reoccurring.

The focus of our work is on the Twitter social media platform (`www.twitter.com`), more precisely on applying learning and classification stategies to learn in the presence of different types of variations of context (*drift*) through time [2,3]. We have used an artificial dataset with Twitter messages that simulates those drift patterns.

2 Background

Twitter stream 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. Besides improving search capabilities, *hashtags* have been identified as having multiple and relevant potentialities, like those described in [4].

Considering the importance of the *hashtag* in Twitter, it is relevant to study the possibility of evaluating message contents in order to predict its *hashtag*. If we can classify a message based on a set of *hashtags*, we are able to suggest an *hashtag* for a given *tweet*. Social networks can be seen as a dynamic and non-stationary environment, in which information is produced by users in a timely order. Time plays a crucial role in Twitter information processing, as past events can give important insights to understand how previously seen information is relevant to improve learning and classification of future unseen and related events. In that sense, learning strategies would be able to learn in dynamic environments and apply innovative strategies to deal with a *"recent memory"* of past events, in order to better identify future and previously unseen ones. There can be several approaches to tackle dynamic environments [5]: instance selection, instance weighting and ensemble learning. A review of concept drift applied to intrusion detection is presented in [6].

3 Proposed Approach

3.1 Twitter classification problem

A Twitter classification problem can be described as a multi-class problem that can be cast as a time series of tweets. It consists of a continuous sequence of instances, in this case, Twitter messages, represented as $\mathcal{X} = \{x_1, \ldots, x_t\}$, where x_1 is the first occurring instance and x_t the latest. Each instance occurs at a time, not necessarily in equally spaced time intervals, and is characterized by a set of features, usually words, $\mathcal{W} = \{w_1, w_2, \ldots, w_{|\mathcal{W}|}\}$. Consequently, instance x_i is denoted as the feature vector $\{w_{i1}, w_{i2}, \ldots, w_{i|W|}\}$. When x_i is a labelled instance it is represented as the pair (x_i, y_i), being $y_i \in \mathcal{Y} = \{y_1, y_2, \ldots, y_{|\mathcal{Y}|}\}$ the class label for instance x_i.

We have used a classification strategy previously introduced in [7], where the Twitter message *hashtag* is used to label the content of the message, which means that y_i represents the *hashtag* that labels the Twitter message x_i.

Notwithstanding being a multi-class problem in its essence, it can be decomposed in multiple binary tasks in a one-against-all binary classification strategy. In this case, a classifier h^t is composed by $|Y|$ binary classifiers.

3.2 Learning Models

In [3] we have studied the impact of longstanding examples in future classification time-windows. The rationale of the presented idea was to store previously seen examples for a period of time regardless the effect they might have as a solo example. Differently from that approach, we are now proposing to choose examples based on the effect they might have individually.

Our baseline model, created for comparison purposes, proposes to store all the information gathered by storing models and combining them as an ensemble. For each time-window, a classifier is trained and stored. When a new collection of documents, in the subsequent time-window, occurs, all the previously trained classifiers are loaded, and the system will classify the newly seen examples. The prediction function of the ensemble, composed by the set of classifiers already created, is a combined function of the outputs of all the considered classifiers. A majority voting strategy where each model participates equally is then put forward. The documents of the previously seen time-windows are not stored in this approach even though the possible learning information is stored along in the classifier trained immediately after it.

We then propose an ensemble learning model, the reinforced model. The main difference is that we define a collection of documents that contains all the classification errors that occur in the time-windows prior to a given moment. The classification errors are considered based on the ensemble classification and not in each model classification output. For each time-window, a classifier is trained with the collection of documents, like in the baseline model, plus the previously introduced error collection and then stored. When a new collection of documents in the subsequent time-window occurs, all the previously trained classifiers are loaded, and will be classified as the newly seen examples participating equally to the final decision of the ensemble. Figure 1 depicts the proposed models.

4 Experimental Setup

4.1 Dataset

The dataset we have defined to evaluate and validate our strategy includes 10 different *hashtags* that represent the different drifts, based on the assumption that they would denote mutually exclusive concepts, like *#realmadrid* and *#android*. By trying to use mutually exclusive concepts we intent to avoid misleading a classifier, as two different *tweets* could represent the same concept.

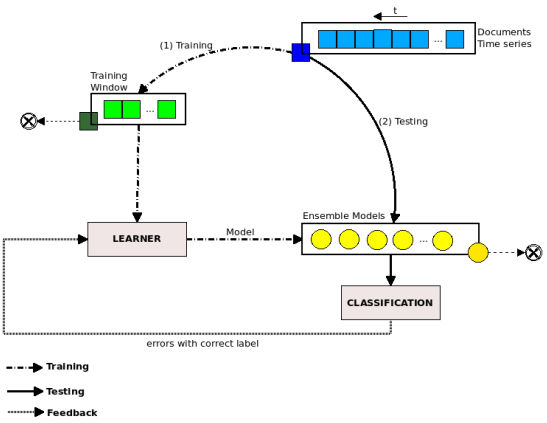


Figure 1: Proposed models

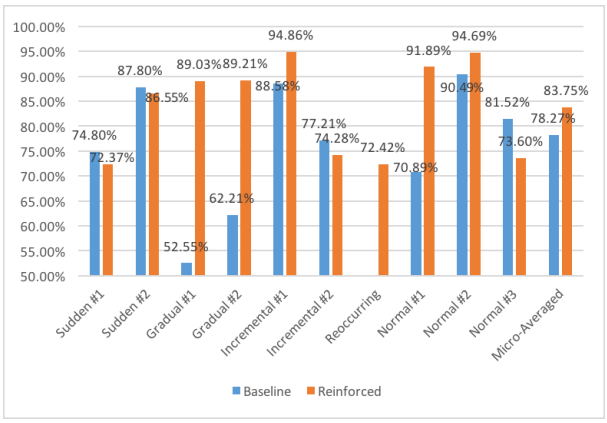


Figure 2: Micro-averaged F1

The Twitter API (`dev.Twitter.com`) was then used to request public *tweets* that contain the defined *hashtags*. The requests have been cared of between December 28, 2014 and January 21, 2015 and *tweets* were only considered if the user language was defined as English. *Tweets* containing no message content besides the *hashtag* were discarded. Finally, the *hashtag* was removed from the message content.

We have simulated the different types of drift by artificially defining timestamps to the previously gathered *tweets*. Time is represented as 100 continuous time windows, in which the frequency of each *hashtag* is altered in order to represent the defined drifts. Each *tweet* is then timestamped so it can belong to one of the time windows we have defined. A profound description of the used dataset can be found in [3]. Our final dataset contains 34.240 *tweets*.

4.2 Representation and Pre-processing

A *tweet* is represented as a vector space model, also known as *Bag of Words*. The collection of features is built as the dictionary of unique terms present in the documents collections. The *hashtag* was removed from the message content in order to be exclusively used as the document label.

High dimensional space can cause computational problems in text-classification problems where a vector with one element for each occurring term in the whole connection is used to represent a document. Also, over-fitting can easily occur which can prevent the classifier to generalize and thus the prediction ability becomes poor. Pre-processing methods were applied in order to reduce feature space. *Stopword removal* was then applied, preventing those non informative words from misleading the classification. *Stemming* method was also applied. Stemming does not alter significantly the information included, but it does avoid feature expansion.

4.3 Learning and Evaluation

The evaluation of our approach was done by the previously described dataset and using the Support Vector Machine (SVM). SVM was used in our experiments to construct the proposed models.

In order to evaluate the binary decision task of the proposed models we defined well-known measures based on the possible outcomes of the classification, such as, error rate ($\frac{FP+FN}{TP+FP+TN+FN}$), recall ($R = \frac{TP}{TP+FN}$), and precision ($P = \frac{TP}{TP+FP}$), as well as combined measures, such as, the van Rijsbergen F_β measure, which combines recall and precision in a single score: $F_\beta = \frac{(\beta^2+1)P \times R}{\beta^2 P+R}$. F_β is mostly used in text classification problems with $\beta = 1$, i.e. F_1, an harmonic average between precision and recall.

5 Experimental Results and Analysis

We evaluate the performance obtained on the Twitter data set using the two approaches described in Section 3, namely the baseline model approach and the reinforced model approach. Figure 2 represents graphically the performance results obtained by classifying the dataset, considering the micro-averaged F_1 measure. Analysing the graph we can observe that globally, and considering the average of the micro-averaged F_1, the storage of the priorly misclassified examples improves the overall classification. This is normal as the learning models are trained with more informative examples and this leads to a better performance. Most classes benefit from storing examples, and we have a significant improve in the average of the micro-averaged F_1, that increases from 78,27% to 83,27%, but some classes, namely *Sudden#1*, *Sudden#2*, *Incremental#2* and *Normal#3* have a worst classification performance. We are confident that this decrease might be explained by the nature of the drift pattern.

As an example, a sudden drift is characterized by an abrupt increase of the frequency of a given class that occur during a period of time, followed by its disappearance. Storing examples that were misclassified, specially the positive ones that appeared firstly and remained misclassified until the classifier identified them as positive, will delude future classifiers, when the drift pattern is no longer represented. Although this is a supposition, that must be validated in future work, we also believe that it might be related to the class, that is the *hashtag* we have chosen to represent it. One of the possible problems that might arise from our approach is to store examples that are not representative of the class.

6 Conclusions and Future Work

We have proposed a method to determine the most relevant examples, by analysing their behaviour when defining separating planes or thresholds between classes. Those examples, deemed better than others, are kept for a longer time-window than the rest. The main idea is to boost the classification performance of learning models by providing additional and significant information.

The results revealed the usefulness of our strategy, as the results improved by 5% in comparing to the baseline approach, considering the average of the micro-averaged F_1. It is also important to conclude that we have shown that retaining informative examples can improve the learners' ability to identify a given class, independently from the drift pattern the class is representing. We do believe that it is problem dependent, even thought it is an important insight in dynamic models, as they are particularly difficult learning scenarios. A special attention must be given to classes that tend to disappear, as retaining examples, in this particular case, for long periods can lead to misclassications.

Our future work will include a more profound study about the longevity of those examples, i.e., for how long is it relevant to retain those examples.

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, Bernardete Ribeiro. Concept Drift Awareness in Twitter Streams. In *Proc. of the 13th Int. Conference on Machine Learning and Applications*, pp. 294-299, 2014.
[3] Joana Costa, Catarina Silva, Mário Antunes, and Bernardete Ribeiro. The Impact of Longstanding Messages in Micro-Blogging Classification. In *Proc. of the International Joint Conference on Neural Networks*, 2015.
[4] M. Zappavigna. Ambient affiliation: A linguistic perspective on Twitter. In *New Media & Society*, vol. 13, no. 5, pp. 788-806, 2011.
[5] A. Tsymbal. The problem of concept drift: definitions and related work. Dept Computer Science, Trinity College Dublin, Tech. Rep., 2004.
[6] J. Kim, P. Bentley, U. Aickelin, J. Greensmith, G. Tedesco, J. Twycross. Immune system approaches to intrusion detection - a review. In *Natural Computing*, vol.6, no.4, pp. 413-466, 2007.
[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*, pp. 226-235, 2013.

Dynamic Recognition of Obstacles for Optimal Robot Navigation

Miguel Fernandes
http://www.di.ubi.pt
Luís A. Alexandre
http://www.di.ubi.pt/~lfbaa

Dep. Informática
Universidade da Beira Interior
6201-001 Covilhã, Portugal
Instituto de Telecomunicações, Torre Norte, Piso 10
Av. Rovisco Pais, 1
1049-001 Lisboa, Portugal

Abstract

Navigation is a well established field with robust algorithms that can work out-of-the-box in systems like ROS. Nonetheless, there are situations in which the current navigation approaches are laking in terms of optimality. Examples arise when too much "safe space" is assigned around a given object that can completely prevent a robot from using a given path and forces the use of an alternative path that can be much longer. In this paper we propose the dynamic adaptation of robot navigation strategies depending on the type of obstacles that are met during navigation. We do this in real time using a convolutional neural network for obstacle recognition and a path planning parameter adjustment depending on the obstacle category. We present experiments illustrating the difference in paths that can be obtained by using the proposed approach versus standard approaches implemented in ROS.

1 Introduction

Robot navigation is an important and active research area since it is one of the fundamental tasks for a mobile robot. In this paper we propose a method that dynamically adapts the cost mapping parameters to the type of obstacle that is present in the robot's path in order to aid the path planner in choosing the best path to take. The objects are recognized using a Convolutional Neural Network (CNN).

In related work regarding navigation of robotic systems, Xin *et al.* [5] proposes a visual navigation system in order to plan a smoother path for the robot to navigate,while taking into account the dimensions of the robot, and being successful in dealing with a dynamic environment.

Courbon *et al.* [1] improves the robustness of localization and the path-following in visual memory-based navigation frameworks with the concepts of short-term and long-term memories.

Menlingui *et al.* [2] develop a new navigation approach by combining Artificial Potential Fields and Interval Type-2 Fuzzy Logic Systems in a omnidrive mobile robot that presents smooth paths that are fast to calculate.

So, although some work has been done in dynamic adjustment of navigation parameters, it has focused on different goals than the ones we are pursuing in this work. Namely, the above works focused improving the navigation by using smoother paths, improved localization and speed in path planning. We are concerned with allowing the robot to be able to choose paths that could be considered as blocked by obstacles and hence allow for eventually using shorter paths than would otherwise be possible.

2 Proposed Method

2.1 Obstacle recognition

We are interested in determining if the obstacles belong to one of two categories: mobile or static objects.

For this, we start by classifying the obstacles in the scene into the 1000 classes of ImageNet. All the animals, transports, and moving objects are mapped into the mobile category. All the others are placed into the static object category. The classification is done on every frame, thus if an object classified as static and then in a subsequent frame is classified as mobile, the algorithm will adjust to the most recent category. To achieve object recognition in real time we take advantage of a previously trained CNN. We use the Extraction model from [4], which is a CNN trained for the ImageNet dataset. It has top-1 validation accuracy of 72.5%.

2.2 Navigation Adjustment

After obtaining the category of the obstacle, the method changes the parameters of the path planner in order to adapt to the obstacle category. The idea is that some objects are "safer" than others. For instance, objects that can move, such as people or animals, require a larger "safety" distance than static objects, like tables or walls.

Navigation in robotic systems is split in two parts, the Path-Planning and the Cost-mapping. A major component of ROS (Robot Operating System) [3] Navigation Stack is the movebase package. It is composed by two planers, two costmaps and a recovery node. One planner and costmap are local, as a dynamic window around the robot, and the other planer and costmap are related to the global map. The cost maps are filled with information from layers, such as obstacle information as lethal points, or an inflationer that inflates points around lethal points in order for the robot to have a safety distance from obstacles.

Our method acts by:

- first recognizing the object closest to the robot in its planned path;
- mapping this object to one of the two categories: static or mobile;
- adjusting the inflation parameters to allow for the robot to pass closer to static than to mobile obstacles.

3 Experiments

In this section we present two experiments that illustrate the benefits of the proposed approach. The first experiment takes place in a virtual environment (Gazebo simulator) and the second on a real environment. Both use a Turtlebot 2 robot equiped with a Kinect camera. We contrast the application of our method to the use of the standard path planner in ROS.

3.1 Experiment 1

In this experiment, the category of the objects is predetermined (no real-time object categorization). We have created a small maze where the robot is placed in the lower left corner and is instructed to navigate to the upper right corner. There are two paths available for this experiment. The shortest path has two coke cans that serve as a static obstacle that, nonetheless, allows the robot to pass. With the regular cost-mapping static inflation method, the robot takes the longer path due to the fact that the inflation ratios affects the coke cans in a way that the cost of using the longer path is smaller than to pass very close to the obstacle. With our layer, the path-planer makes the robot use the shortest path although it passes very close to the static obstacles.

3.2 Experiment 2

In this experiment we have a setup in our lab that contains two possible paths from point A to B. The first one is shorter but forces the robot to pass very close to an obstacle (under a tripod) – see Fig. 3.

The standard setup for navigation in the ROS stack does not allow the robot to use this path because the safety distance that is used by the planner forces the robot to consider the path as blocked (Fig. 4).

Our method classifies the type of obstacle in the static category and hence assigns it a low probability of motion so allows for a closer approximation of the robot to the obstacle and makes the shorter path usable (Fig. 5). In this experiment, the category of objects is recognized in real-time using a CNN. The obstacle is recognized as a tripod when the robot is about 1 meter away from it. The recognition code runs on the GPU (Titan X) and takes around 10 ms to recognize the object in each frame,

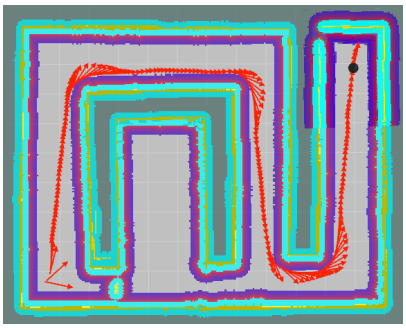


Figure 1: Experiment 1. Map showing the path used by the robot when working with the standard ROS configuration: when faced with the static obstacles in the shortest path, the robot had to choose the longest path.

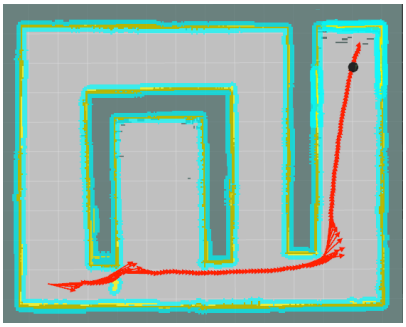


Figure 2: Experiment 1. Map showing the path used by the robot when working with our proposed approach: by recognizing that the obstacle is static it could adjust its parameters to allow for passing closer to the obstacle, making the shorter path available for navigation and saving time.

allowing for real-time recognition. The same code running on CPU takes 5s per frame. We took advantage also of the highly optimized CuDNN library for deep learning.

4 Conclusions

Navigation in unstructured environments is challenging and in this paper we improve the navigation stack of ROS by allowing the recognition of the type of obstacle present in the scene and automatically adjusting the navigation parameters to allow a reduction or increase in the safety space that is considered around obstacles.

We show that this allows a robot to use paths that were previously considered as blocked and can thus improve its navigation performance. We compare in two experiments the results of our method against the standard ROS navigation stack.

Future work will be done to use more object categories and have a finner adjustment of the path planner parameters.

5 Acknowledgments

This work was supported by National Funding from the FCT - Fundação para a Ciência e a Tecnologia, through the UID/EEA/50008/2013 project.

The GTX Titan X used in this research was donated by the NVIDIA Corporation.

References

[1] J. Courbon, H. Korrapati, and Y. Mezouar. Adaptive visual memory for mobile robot navigation in dynamic environment. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 873–878, June 2012. doi: 10.1109/IVS.2012.6232166.

[2] A. Melingui, T. Chettibi, R. Merzouki, and J. B. Mbede. Adaptive navigation of an omni-drive autonomous mobile robot in unstructured dynamic environments. In *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1924–1929, Dec 2013. doi: 10.1109/ROBIO.2013.6739750.

[3] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully B. Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. ROS: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

[4] Joseph Redmon. Darknet: Open source neural networks in C. `http://pjreddie.com/darknet/`, 2013–2016.

[5] J. Xin, X. L. Jiao, Y. Yang, and D. Liu. Visual navigation for mobile robot with kinect camera in dynamic environment. In *2016 35th Chinese Control Conference (CCC)*, pages 4757–4764, July 2016. doi: 10.1109/ChiCC.2016.7554091.

Figure 3: Experiment 2 setup showing the lab, the robot on the starting position and the path with obstacles (two computer boxes and a tripod). The photo was taken standing on the end position.

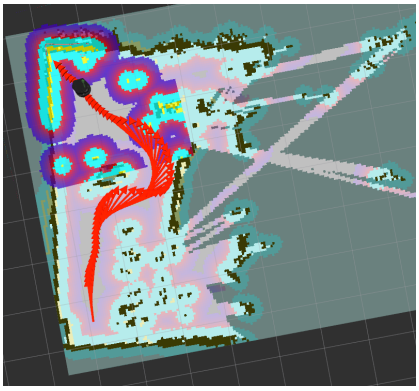


Figure 4: Experiment 2. With the standard ROS approach, the robot has to take the longer path when faced with the tripod obstacle.

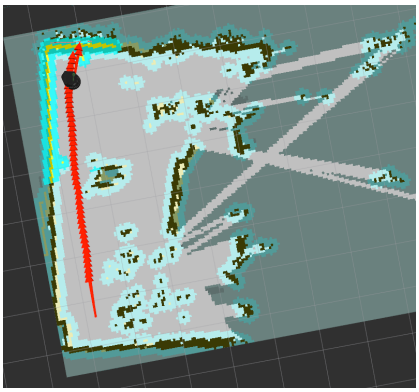


Figure 5: Experiment 2. With our proposal, the robot recognizes the tripod as a static obstacle and adjusts the safety margin around the object, enabling the use of the shortest path.

Initial validation of online ECG signal segmentation

Tiago Magalhães
tiagomagalhaes@ua.pt
José Maria Fernandes
jfernan@ua.pt
Ilídio Castro Oliveira
ico@ua.pt
Susana Brás
susana.bras@ua.pt

IEETA, DETI
Universidade de Aveiro
Aveiro, Portugal

Abstract

Acquisition of bio-signals in the field (such as ECG signal) is prone to artifacts and noise. This may jeopardize relevant information and mislead automatic analysis algorithms. In this paper we propose an algorithm for online ECG noise detection based on RR intervals analysis, with an accuracy above 99%. The current implementation runs on smartphones and allows for an immediate assessment on whether the signal is worth to transmit, raising opportunities to enhance cyber-physical systems development.

1 Introduction

In real world conditions, data collected from biomedical sensors are prone to artifacts and noise [1, 8]. As the duration of the collected signals makes the visual inspection difficult to perform, automatic systems were developed in order to help the therapist in their signal evaluation [3, 4, 5, 9]. In the specific case of electrocardiogram (ECG), the collection of data on duty activities and daily life scenarios allows the inference and quantification of ECG alterations associated with, *e.g.* pathologies, emotions or often external variables. However, in many cases the ECG had intervals with noise that lead to misleading interpretations in automatic algorithms. Examples are the extraction of respiration trend, identification of R-peaks, identification of fatigue levels, etc.

In literature, several algorithms are presented for noise reduction on ECG [6, 7, 8]. However, this paper presents an algorithm to identify intervals where noise is present. Two algorithms had already been presented. In Bras *et al*. [1] alterations over consecutive windows of ECG were analysed with the assumption that it should not be present abrupt alteration when the ECG signal is noise free. The implemented rules are over the ECG signal, and use measures that compromise the computational efficiency. Varon *et al*. [8] present in their paper an algorithm based on the autocorrelation function over ECG windows. They present good results nevertheless the algorithm is not online, needing the whole ECG signal to perform the algorithm. So, in our system, the disadvantages of both algorithms prevent their implementation.

Considering all these arguments, in this paper, we present an algorithm for online ECG noise detection. The idea was to use the RR interval, which allowed a computationally efficient algorithm, since the RR-intervals were calculated by the used ECG wearable system.

2 Dataset

This dataset is composed by 19 hours (7 signals with 2h40 each, approximately) corresponding to a clean ECG signal, collected using a simulator, with synthetic noise added (as shown in Fig. 1). To gather the values from the simulator we used VitalJacket [2], which provides an ECG signal with a sampling frequency of 500 Hz. The collecting protocol consisted in gradually increasing and decreasing the heart rate (HR) with 20 minutes interval. In total, it is composed by 8 HR steps in the following order: 60, 80, 100, 120, 140, 120, 100, 80 beats per minute (bpm).

The noise values were randomly generated using a uniform discrete distribution between the minimum and maximum values (117 and 159, respectively) of the collected signal. This allows us to have a controlled signal with specific noisy zones to test the algorithm behaviour in different situations. In the first signal it was inserted 10 seconds of noise during the transition between 60 and 80 bpm, omitting it (Fig. 1). The second signal contains 90 seconds of synthetic noise in the middle of the 80 HR step

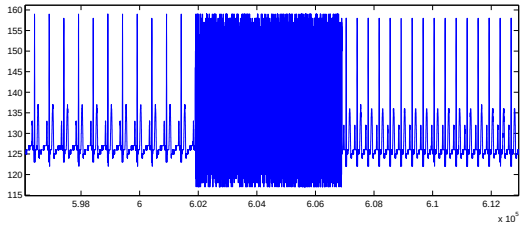


Figure 1: 10 seconds of artificial noise introduced between the 60 and 80 bpm omitting the transition.

and, in the third one, it was introduced 15 minutes in the 140 HR step covering a large part of this segment and hiding the transition between 140 and 120 bpm. The fourth signal begins with 20 seconds of noise and only after it there is the clean signal. The transition between 120 and 100 bpm, in the fifth signal, was replaced by 90 seconds of noise and, in the sixth signal, 10 seconds was replaced by noise in the middle of the second time of the 80 bpm step. At last, the seventh ECG signal contains all the noise locations and lengths outlined above.

3 Method

The algorithm has been developed taking two assumptions into consideration. At first, that there is no abrupt change on the HR, i.e. the HR might increase and decrease gradually but not instantly. Therefore, the algorithm analyses the ECG RR intervals (temporal difference between two consecutive R peaks) and makes a decision based on a superior and inferior established thresholds.

The other assumption is that the algorithm has a limited input historic, as it is intended to be, used for online processing of ECG. Currently, it only takes the actual RR interval value and the last three RR values into consideration.

Based on the assumptions outlined before, we have developed the following algorithm. Firstly, it looks for 3 (heart) beats with 30 bpm or less – considered ECG signal with noise absence – to establish a RR value of reference and start the noise detection. The RR reference value corresponds to the average of the last three stored RR values. This value is used to calculate the respective superior and inferior ranges. Due to the non-linear relation between RR and HR these ranges must be computed in each interaction in order to use thresholds with 30 bpm of range.

Both ranges indicate the acceptable deviation ($\pm$30), from the reference value, in which the RR value that is being analyzed should be in order to be marked as signal. In this case, the actual RR value is replaced by the average of the last three RR values and a set-value based on the actual RR interval and its distance to the reference RR value. Thus, it will reflect the oscillation trending of the RR values smoothly.

If a RR value is out of the limits it is considered as noise and the algorithm returns to the first phase, which looks for a valid ECG signal.

4 Tests

In order to perform the tests, the ECG signals were pre-processed to obtain the RR intervals using an algorithm based on the Pan and Tompkins [3, 5] algorithm. The resulting files were used as input for the test program. This program simulates the incoming RR values of the real system

Signal #	Fs (Hz)	Duration (h:m:s)	Noise added (%)	Noise detection (%)	Noise sensitivity (%)	Noise specificity (%)	Accuracy (%)
1	500	02:41:29	0.10	0.12	94.8800	99.9737	99.9684
2	500	02:41:29	0.93	0.93	99.1289	99.9866	99.9786
3	500	02:41:29	9.29	9.07	97.6067	99.9954	99.7735
4	500	02:41:29	0.21	0.14	52.8300	99.9709	99.8735
5	500	02:41:29	0.93	0.94	99.1289	99.9759	99.9680
6	500	02:41:29	0.10	0.13	98.7000	99.9694	99.9681
7	500	02:41:29	11.56	11.34	97.0371	99.8560	99.5302

Table 1: Noise detection results obtained using the proposed algorithm.

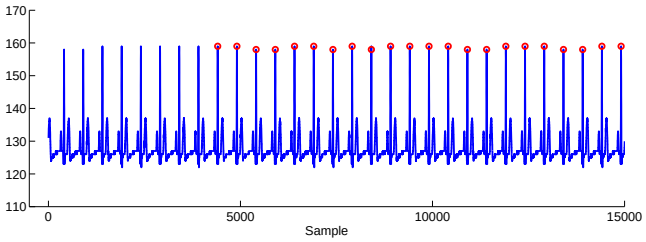


Figure 2: Beginning of the signal 1. ECG signal in blue and R-peaks in red.

implementation of the algorithm. In the real scenario, the VitalJacket collects the ECG and provides the RR values stored into the data structure and consumed by the algorithm.

5 Results and discussion

Table 1 shows the results obtained using the described algorithm where it is possible to assess the performance of the algorithm in terms of sensitivity, which gives the probability of the noise be detected as noise by the algorithm, specificity, which indicates the probability of the algorithm not identify noise and actually the evaluated sample is not noise, and the accuracy, which evaluates the degree of true evaluations.

Except the signal 4, the sensitivity values are greater or equal than 94.88% showing good capacity of the algorithm to detect noise. The test formulation in the signal 4 leads to a weak result of a sensitivity of 52.83%. In this test, it was inserted 20 seconds of noise at the beginning of the signal. As shown in the Fig. 2, the Pan Tompkins' algorithm needs about 8 beats to start providing the RR intervals. With noise at the beginning of the signal (Fig. 3), few RR values are retrieved from the noisy signal and, additionally, our algorithm can not provide a result for the first three RR values, once it needs a historic of the last 3 RR values. Thus, the algorithm detected few noisy samples obtaining a poor sensitivity value. However, all signals have a good accuracy with an average of 99.87%.

Another algorithm behaviour is that in presence of three similar RR values in the noise zone produces a misleading and, consequently, false negatives. Additionally, when the algorithm detects true signal, 4 beats (3 RR values) are lost (Fig. 4) resulting in false positives.

6 Conclusion

In this paper we propose an algorithm for online noise detection for ECG signal. The results indicate that the algorithm is capable to accomplish its goal with an accuracy greater than 99%. The current implementation of this algorithm on a smartphone will allow the system reacting to noise

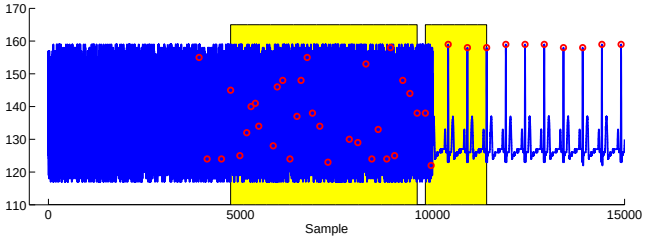


Figure 3: Beginning of the signal 4 with 20 seconds of noise added. The noise area detected in yellow.

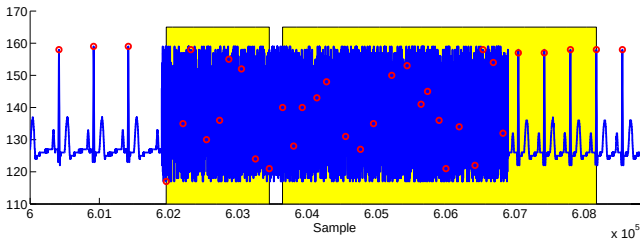


Figure 4: ECG signal 1 with 10 seconds of noise added.

detection efficiently, such as stop sending the ECG signal or turn off Wi-Fi resources, enhancing, for example, the energy consumption, as well as the online labeling of the signal for further analysis.

7 Acknowledgments

This work was supported by the European Regional Development Fund (FEDER) and FSE through the COMPETE programme and by the Portuguese Government through FCT - Foundation for Science and Technology, in the scope of the projects UID/CEC/00127/2013 (IEETA/UA), CMUPERI/ FIA/ 0031/2013 (VR2Market), and PTDC/EEI-SII/6608/2014. S. Brás acknowledges the Postdoc Grant from FCT, ref. SFRH/BPD/92342.

References

[1] Susana Brás, Nuno Ferreira, and João Paulo da Silva Cunha. ECG Artefact Detection Algorithm-An Algorithm to Improve Long-term ECG Analysis. In *BIOSIGNALS*, pages 329–333, 2012.

[2] J. P. S. Cunha, B. Cunha, A. S. Pereira, W. Xavier, N. Ferreira, and L. Meireles. Vital-Jacket®: A wearable wireless vital signs monitor for patients' mobility in cardiology and sports. *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*, pages 1–2, 2010.

[3] Patrick S. Hamilton and Willis J. Tompkins. Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database. *IEEE Transactions on Biomedical Engineering*, BME-33 (12):1157–1165, 1986.

[4] Juan Pablo Martínez, Rute Almeida, Salvador Olmos, Ana Paula Rocha, and Pablo Laguna. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE transactions on bio-medical engineering*, 51(4):570–81, 2004.

[5] Jiapu Pan and Willis J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32 (3):230–236, mar 1985.

[6] R. Sameni, M.B. Shamsollahi, C. Jutten, and G.D. Clifford. A Nonlinear Bayesian Filtering Framework for ECG Denoising. *IEEE Transactions on Biomedical Engineering*, 54(12):2172–2185, 2007.

[7] K. Sharmila, E. Hari Krishna, Komalla Nagarjuna Reddy, and K. Ashoka Reddy. Application of multiscale principal component analysis (MSPCA) for enhancement of ECG signals. In *2011 IEEE International Instrumentation and Measurement Technology Conference*, pages 1–5. IEEE, 2011.

[8] Carolina Varon, Dries Testelmans, Bertien Buyse, Johan A K Suykens, and Sabine Van Huffel. Robust artefact detection in long-term ECG recordings based on autocorrelation function similarity and percentile analysis, 2012.

[9] J A Vila, Y Gang, J M Rodriguez Presedo, M Fernández-Delgado, S Barro, and M Malik. A new approach for TU complex characterization. *IEEE transactions on bio-medical engineering*, 47(6):764–72, 2000.

Anomaly-based intrusion detection using application-specific traffic profiles

Hassan Alizadeh
hassan.alizadeh@ua.pt
IEETA, University of Aveiro

André Zúquete
andre.zuquete@ua.pt
Department of Electronics, Telecommunications and Informatics, IEETA, University of Aveiro

Abstract

We address the problem of detecting intrusions in (known) applications when their traffic exhibits anomalies. To do so, we need to: (1) bind traffic to applications; (2) have per-application traffic profiles; and (3) detect deviations from profiles given a set of traffic samples. This work was mainly devoted to the last two topics. Upon an initial survey on traffic classification techniques, we propose the use of two different kinds of Gaussian Mixture Model learning approaches to build per-application profiles and we tested them against public datasets where the source of each traffic flow is provided. Experimental results indicate that the proposed approaches are effective.

1 Introduction

Along with the growing number of applications and end-users, online network attacks and advanced generations of malware have continuously proliferated. Since most intrusions result in network activities, an effective strategy to detect these intrusions is to have robust network traffic inspecting techniques. Detection of intrusions from traffic inspection, as part of a Network Intrusion Detection System (NIDS), can be performed by matching observed traffic against a pre-configured set of (well-known) intrusion signatures (signature-based NIDS), or by noticing deviations from pre-defined models (profiles) describing normal traffic (anomaly-based NIDS). Unlike signature-based NIDS (SNIDS), anomaly-based NIDS (ANIDS) may detect new types of intrusions (unknown) and tackle the so-called zero-day attacks. Due to the advantages of ANIDS over SNIDS paradigms regarding zero-day attacks, most existing research studies have focused on ANIDS.

Much of the history of ANIDS has focused on inspecting aggregated network traffic to detect deviations from normal behavior with no knowledge of the responsible applications. Such systems fail to detect intrusions in applications whenever their abnormal traffic fits into the network normality profiles. For example, Thunderbird uses mail-related protocols (SMTP, POP, IMAP, etc.); but under the influence of a sophisticated intrusion, it may generate some other kinds of protocols, such as well-formed HTTP. Although such intrusion causes Thunderbird to deviate from its normal behavior, this deviation may be within the normal behavior of the network. To handle such situations, the process of anomaly detection should be separately applied on each application's traffic. To do so, an ANIDS requires to: (i) identify the claimed application and (ii) have per-application traffic profiles. Regarding the first requirement the architectures presented in [1, 2] provide a binding between network traffic and source application.

Regarding the second requirement, in [3] we investigated traffic classification strategies, within a taxonomy framework, and discussed how the referred approach could help us to manage applications' traffic profiles. As results of this study, we proposed the use of two different kinds of Gaussian Mixture Model (GMM) learning approaches, namely universal background model (UBM-GMM) and unsupervised learning of GMM (uGMM), to build applications' traffic profiles from their normal flows' features and thereby form detection systems.

The rest of the paper is organized as follows. Section 2 describes the form of GMM and its parameterization. In section 3, we briefly describe the use of UBM-GMM in detecting anomaly in (known) applications and present the details of experimental setups and evaluation results. Section 4 briefly describes uGMM and presents experimental results. Conclusions and possible future work are discussed in Section 5.

2 Gaussian Mixture Model (GMM)

The underlying distribution of flows' feature values extracted from an application *a* can be modelled by a Gaussian mixture density. The Gaussian mixture density is a weighted linear combination of *M* component Gaussian densities. For a D-dimensional feature vector *x*, the mixture density for application *a* is defined as:

$$p(x \mid a, \theta^a) = \sum_{m=1}^{M} \omega_m^a \, p(x \mid a, \theta_m^a) \quad (1)$$

where $\theta^a = \{\theta_1^a \dots \theta_M^a, \omega_1^a \dots \omega_M^a\}$ is the set of all parameters of the model, ω_m^a are the mixture weights (satisfying the constraints $\omega_m^a > 0$ and $\sum_{m=1}^{M} \omega_m^a = 1$) and θ_m^a are a set of parameters, namely a mean $D \times 1$ vector, μ_m^a , and a $D \times D$ covariance matrix, Σ_m^a , that defines the m^{th} component unimodal Gaussian density function as follow:

$$p(x \mid a, \theta_m^a) = \frac{1}{(2\pi)^{D/2} \mid \Sigma_m^a \mid^{1/2}} e^{-\frac{1}{2}(x-\mu_m^a)'(\Sigma_m^a)^{-1}(x-\mu_m^a)} \quad (2)$$

The parameters of an application's *a*'s density model are collectively denoted as $\theta^a = \{\omega_m^a, \mu_m^a, \Sigma_m^a\}$ where $m = 1, \dots, M$. These parameters are estimated via Expectation-Maximization (EM) algorithm. The EM is an iterative procedure that begins with an initial model θ_a, to estimate a new model $\hat{\theta}_a$ such that $p\left(x \middle| a, \hat{\theta}_a\right) > p(x|a, \theta_a)$. For the next iteration, the new model becomes the initial model and the process is repeated until some convergence threshold is met.

3 Universal background model

Given a traffic flow sample, *T*, along with its claimant's application, *a*,

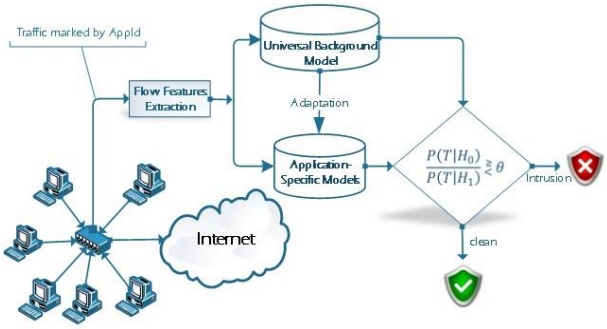


Figure 1: Application of the UBM to detect abnormal traffic samples

we define the problem of abnormality detection in traffic of specific application as the determination of whether or not *T* was generated by *a* [4]. This problem can be interpreted as a basic hypothesis test between H_0 (*T* was generated by *a*) and H_1 (*T* was not generated by *a*) as the null and alternative hypothesis, respectively. The optimal decision test between these two hypotheses can be taken by a likelihood-ratio test:

$$\varphi(T \mid H_0) = \frac{p(T \mid H_0)}{p(T \mid H_1)} \begin{cases} \geq \theta & \text{accept } H_0 \\ < \theta & \text{accept } H_1 \text{ or reject } H_0 \end{cases} \quad (3)$$

where $p(T|H_i)$, $i \in \{0,1\}$ is the *likelihood* of observing sample *T* under hypothesis H_i and θ is a user-defined decision threshold for accepting or rejecting the claim. θ can be obtained based on various criteria (see Sec. 3.1.2). H_0 should characterize the hypothesized specific application, whereas, H_1 should be able to model *all the alternatives to the hypothesized specific application*. Figure 1 shows a block diagram of the proposed system.

A model describing H_1 is known as Universal Background Model (UBM), which can be built by fitting a GMM to the underlying distribution of flow-level features of all applications. An application-specific model can be obtained using its traffic samples by the adaptation of the UBM's parameters through Maximum A Posteriori (MAP) estimation.

3.1 Experimental Methodology

3.1.1 Flow Definition and Feature Extraction

Experiments were conducted on the "measurement" dataset[1] where the source of each traffic flow is provided. We define a flow by a set of bidirectional consecutive packets traveling between two endpoints (defined by srcIP:srcPort and dstIP:dstPort) through a TCP connection started by a 3-way handshake (SYN, SYN-ACK and ACK) and terminated by either observing FIN/RST packets or no packets is seen for a timeout of 60s (whichever comes first). The first SYN packet seen in a flow determines the forward direction and the source endpoint is assigned to client. We considered only TCP flows that have at least one packet in each direction and contain at least one non-zero payload packet.

A flow associated with a particular application can be described by a number of statistical properties, or features, parameterizing its behaviour. We extracted the above defined bidirectional flow objects and calculated a total of 39 statistical feature values from the trace file. Such features can be computed directly from some parts of packet headers (such as average packet size, total transferred bytes) as well as indirectly from the time when a packet arrives (i.e. inter-packet timings such as min/max/average/variance packet inter-arrival time, flow duration).

3.1.2 Experimental Setup and Evaluation Metrics

TCP flows generated by a set of 9 applications were randomly divided in three disjoint sets: training, evaluation and test. For both evaluation and test stages, we simulated the infected traffic of a specific application using the normal traffic of other applications, claiming its identity.

For each claim, the system performs a likelihood test and outputs a normality score reflecting how well the probe flow object matches the claimed identity. These scores were generated for the evaluation set in order to find an optimal value for the threshold (θ). This threshold determines if a specific application is clean or not. The threshold can be defined in two ways: 1) *Global Threshold (GT)*, where a single identical threshold is defined for all the classes; 2) *Class-Specific Threshold (CST)*, which employs a different threshold for each class (application).

In this study, the threshold is determined based on the *Equal Error rate (EER)* criterion, i.e., by the operating point where the *False Rejection Rate (FRR)* is equal to *False Acceptance Rate (FAR)*. False Acceptance happens when a traffic flow generated by an infected application is miss-detected as genuine. False Rejection happens when traffic produced by a clean application is considered infected. We then used the set threshold in the test, from which the FAR and FRR are calculated. We report Half Total Error Rate (HTER), which takes the average of FAR and FRR as a single measurement. The FAR, FRR and HTER range from 0% (best) to 100% (worst).

3.1.3 Experimental Results

We compared the evaluation results of overall HTER in both Evaluation and Test subsets for different number of mixture components in order to obtain an optimal number of components for UBM-GMM. We observed that a mixture of 16 components with the use of CST technique yielded a lower overall HTER in both subsets (2.44% and 2.45% for Evaluation and Test subsets, respectively). Table 1 illustrates the performance results of CST technique for each single applications where a mixture of 16 components was used. Promising results (HTER < 8% for all applications in both Evaluation and Test sets) indicates the effectiveness of UBM-GMM profiling approach.

Table 1: Per-applications performance results obtained by the CST

Applications	Evaluation Set			Test set		
	FAR(%)	FRR(%)	EER(%)	FAR(%)	FRR(%)	HTER(%)
wpc55agv2	0.79	0.78	0.78	0.85	1.00	0.92
eMule	2.49	2.5	2.49	2.64	1.78	2.21
Azureus	2.73	2.73	2.73	2.52	2.87	2.69
Internet explorer	2.94	2.94	2.94	2.95	3.18	3.06
MS Outlook Exp	0.16	0.16	0.16	0.21	0.85	0.53
FilePlanet	2.34	2.35	2.35	2.37	2.03	2.20
Skype	4.06	4.06	4.06	4.48	3.51	3.99
MSN Messenger	4.95	5.26	5.11	4.85	9.65	7.25
Limewire	6.64	6.38	6.51	6.37	9.10	7.74

[1] http://www.crysys.hu/~szabog/index_files/measurement.tar

4 Unsupervised Learning of GMM

The main challenge of the UBM approach is that, due to the use of the basic EM algorithm in training UBM, the number of components has to be known a-priori, which causes the difficulty of building a whole set of candidate models for selecting an optimal model. To deal with this problem, we proposed [5] the use of an unsupervised learning of GMM (uGMM) presented in [6], which automatically selects the optimal number of components for a given data. The method facilitates a seamless integration of parameters estimation and model selection in a single algorithm by implementing a variant of the EM algorithm, termed component-wise EM (CEM) [7], that aims at minimizing a minimum message length (MML) like criterion as a cost function.

4.1 Experimental Methodology

Experimental methodology used for uGMM is almost similar to UBM approach. However, a likelihood test replaced the likelihood-ratio test, since universal model is not trained. The proposed approach was evaluated using experiments conducted in UNIBS traces[2]. Table 2 presents the best performance results for each group of applications which was obtained by CST technique.

Table 2: Per-applications performance results obtained by the CST for different applications in UNIBS traces

Applications	Evaluation Set			Test set		
	FAR	FRR	EER	FAR	FRR	HTER
MAILS (Apple Mail, Thunderbird)	2.24	2.23	2.24	2.50	2.30	2.40
P2P(Transmission, eMule, Bittorrent)	6.69	6.72	6.70	6.31	8.64	7.47
Browsers (Safari, Firefox, Opera)	8.67	8.67	8.67	9.05	8.98	9.01

5 Conclusions

Motivated by the goal of detecting intrusions in (known) applications using their traffic samples, two application-level intrusion detection frameworks were presented, based on UBM-GMM and uGMM algorithms. These were used to build robust models for genuine individual classes. The positive results indicate the effectiveness of the frameworks.

In an ongoing work, the effectiveness of different number of first initial packets of flows was explored and evaluated in order to provide an efficient and timely application-aware traffic anomaly detection system.

The main direction for our future work is to extend our framework in order to be applicable in an online non-stationary environment, where both class evolution and concept drift are available in time.

References

[1] A. Zúquete, P. Correia, and H. Alizadeh, "Packet tagging system for enhanced traffic profiling," in *5th IEEE International Conference on Internet Multimedia Systems Architecture and Application*, Bangalore, Karnataka, pp. 1-6, 2011.

[2] A. Zúquete and M. Rocha, "Identification of source applications for enhanced traffic analysis and anomaly detection," in *IEEE International Conference on Communications*, Ottawa, ON, pp. 6694-6698, 2012.

[3] H. Alizadeh and A. Zúquete, "Traffic classification for managing Applications' networking profiles," *Security and Communication Networks*, vol. 9, pp. 2557-2575, 2016. 10.1002/sec.1516.

[4] H. Alizadeh, S. Khoshrou, and A. Zúquete, "Application-Specific Traffic Anomaly Detection Using Universal Background Model," in *Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics*, San Antonio, Texas, USA, pp. 11-17, 2015.

[5] H. Alizadeh, A. Khoshrou, and A. Zuquete, "Traffic classification and verification using unsupervised learning of Gaussian Mixture Models," in *IEEE International Workshop on Measurements & Networking*, pp. 1-6, 2015.

[6] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 381-396, 2002. 10.1109/34.990138.

[7] G. Celeux, S. Chretien, F. Forbes, and A. Mkhadri, "A Component-Wise EM Algorithm for Mixtures," *Journal of Computational and Graphical Statistics*, vol. 10, pp. 697-712, 2001. 10.1198/106186001317243403.

[2] http://www.ing.unibs.it/ntw/tools/traces/

Mobile Application in the Executive Function Assessment of Parkinson's Disease

Tiago Fonseca[1], Sofia Pires[1],
{a21220551, a21220549} @alunos.isec.pt
Verónica Vasconcelos [1,2], Emilia Bigotte [1,3]
{veronica, ebigotte} @isec.pt

[1] Coimbra Institute of Engineering, Polytechnic Institute of Coimbra
[2] INESC TEC, OPorto
[3] CASPAE, Coimbra

Abstract

In this paper is presented a project that aims to stimulate and evaluate the executive function in patients with Parkinson´s Disease. This project is being developed in partnership with the private social solidarity institution CASPAE, and with the Coimbra Hospital and Universitary Centre (CHUC). This project is a response to a specific need evidenced by health professionals of the Neurology Service at CHUC, during medical appointments with Parkinson's patients. One of the more commonly tests used in diagnosis and follow-up Parkinson's patients is called Trail Making Test (TMT). The fact of the test be performed on paper raises some issues for health professionals. These problems led to the need to convert the TMT to a digital version, using the Android OS.

The developed application allows a simpler and organized test. It has available two operating modes: the "Appointment Mode" that allows health professionals to do the test and save the results more effectively, and the "Train Mode", which allows patients to train TMT on your smartphone or tablet.

1 Introduction

The Parkinson disease (PD) is a degenerative, chronic and progressive disease of the central nervous system. This disease affects more frequently the elderly people [1]. Although the disease mainly affects motor skills, there are some visible changes in his executive function, affecting his attention, his ability to problem solving, and the ability of sequencing [2].

Few applications in the market allow the evaluation and stimulating of the executive function of a patient. In collaboration with CASPAE and CHUC, we developed an application that allows the assessment of the executive function in an easier and simpler way, and adapted to the new technologies.

The application was developed using the Android operating system, which allows its use in common smartphones and tablets. With the developed application, the user can easily perform the test and have access to results automatically. Besides the possibility to perform the test, the application also allows saving the results of the test, making easier compare the results of tests performed previously.

The application can be used in two different ways: It can be used by health professionals during the medical appointment of the patient, to help diagnose or evaluate the progression of the disease, or it can be used by the patient to train his executive function at home, because the application generates random tests.

2 Trail Making Test

The TMT is very popular due to their precision, simplicity and fast implementation. The TMT is very popular due to their precision, simplicity and fast execution. Is given the patient a sheet of paper with various sequences of numbers or letters scattered randomly. The patient should be able to identify the correct sequence in the shortest possible time and with the fewer number of errors [3].

Figure 1 shows an example of the test performed by a patient during a medical appointment. The test consists of two parts (Part A and Part B). In Part A, the patient has to draw a path following the numbers from 1 to 25, in ascending order. In Part B, he must draw a path following the numbers 1 to 13 and the letters A to M alternatively and once again in ascending order, e.g. 1 – A – 2 – B – 3 - C. Each part has also an example, so the patient can understand the rules of the test. For Part A the example has the numbers from 1 to 8 and for Part B the example has the numbers from 1 to 4 and the letters from A to D. When the patient begins to do the test, the doctor starts a timer to record the test duration, and simultaneously observe how many mistakes the patient makes during the test. When the patient finishes the doctor stops the timer and records the results .

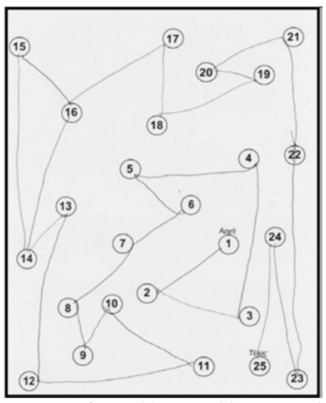


Figure 1: TMT Part A performed by a Parkinson's patient during a medical consultation.

The time that the patient takes to solve the sequences of the test depends, essentially, from the following factors [5]:

- Age affects both part A and part B of test; older people take more time to perform the sequence.
- Education has influence in part B because sequences with numbers and letters are introduced. People with little education have more difficulty in performing the test.
- Patients with some visual disturbance have worse outcomes than patients without any vision problems.

2.1 Problems Observed in Traditional TMT

The realization of the test in paper has some drawbacks:

- The tests available are always the same. So, if a patient makes the same test more than one time, the results may be biased by the memory of the last test, therefore the results do not have any clinical significance;
- Sometimes, the test resolution is confusing due to overlapping lines drawn by the patient to link the numbers and / or letters, making difficult to count the errors on the part of the physician;
- The paper solution does not allow a history of the patient's results. This information can be valuable to the follow-up of the disease;
- The time clocked by the physician is never 100% correct because it is controlled visually, and can exist differences between the real time and the measured time.

3 Proposed Solution

The proposed solution is to adapt the TMT to new technologies in order to allow the patient to be able to make this test as many times as you want on your own smartphone or tablet. And on the same side, solving the problems described above.

- The time of execution is more reliable in the developed application. The patient clicks a button to start a timer that stops when the test ends, allowing the record of the exact time of duration of the test;
- During the test, numbers or letters correctly marked turn green and the wrong selection turn red. So it is possible to immediately see the errors and the sequence isn't so confused because isn't used pen and paper to draw the right path;
- At the end of the test, when all the numbers and letters of the sequence are green, meaning the test finished and the sequence is correct, the number of errors are counted as well as the duration of

the performed test. Every time a button turns red, a penalty is added to the final results. After the patient finishes the test a message will appear with the results: time and number of errors.

- In all the tasks, the distribution of letters and numbers is done randomly. Therefore, patients can make the test many times, without the possibility of memorizing the sequences.

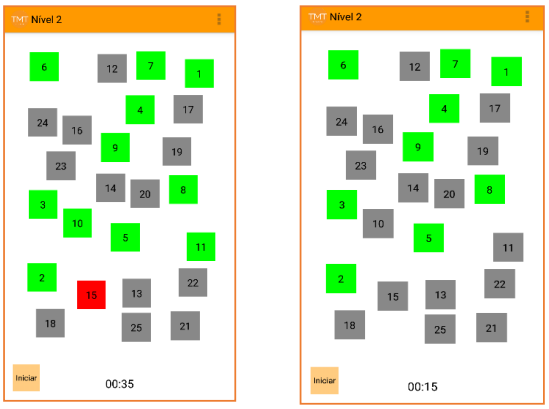


Figure 2: Level 2 of the test. Level On the left: TMT Part A without errors (only green squares); On the right: TMT Part A with one error (red square)

The installation of the application consists on the download of the file from the Play Store (Android System App Store). The smartphone/tablet will automatically install the app in few seconds. The application has two operating modes: the "Appointment Mode" to health professionals to use during medical appointments to in the diagnosis or follow-up of the disease, and the "Train Mode" to be used by the patient whenever and wherever he wants, on his own smartphone or tablet. The application has four levels on both modes [6]. Examples of resolved tests for the Level 2 are shown in Figure 2. In the "Appointment Mode" the patient has to fill a small form with the identification, name and education level, before start doing the test. The age and education level are important data since can affect the test results. In the "Train Mode" is necessary to insert a username and a password before starting the test. When the test ends, the results are shown, allowing the patient or family to analyze them. Since each test needs a user identification is possible to use the same the application by more than one person in the same mobile device, and perform the test without interfering with other users' data.

Both application modes have the possibility of save the tests results. The stored data can be valuable for the medical team that follows the patient and to the patient. The relational model of the database used is present in Figure 3.

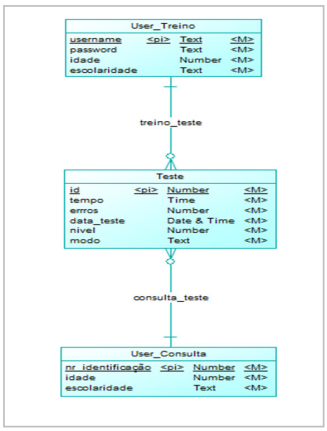


Figure 3: Conceptual Model of the local database, of both modes of the application.

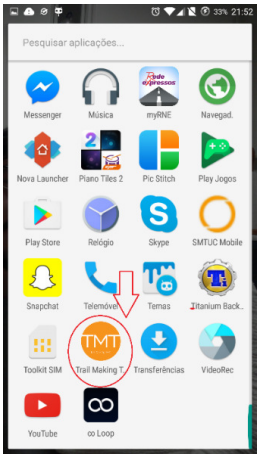


Figure 4: Logo of the developed application in smartphone screen

4 Conclusions and Future Work

An application for the Android platform was developed to allow stimulation and assessment of executive function in Parkinson's disease. This application implements the Trail Making Test to be used by physicians and patients in a simple and useful way.

In the future we would like to test the application on more patients with Parkinson's disease to compare the results with those obtained with the paper test. The authors also have intention to create a different type of level that has never been used before in TMT, for patients with low level of education, using logical sequences, e.g. seed - tree - flower – fruit. In order to make the application available for community, we intend to put it in the Play Store. The logo developed for the application can be seen in Figure 4.

Acknowledgments

The authors thank Dr. Cristina Januário and her medical team from Neurology Service of Coimbra Hospital and University Centre, for their medical knowledge and assistance. We also thank to the students Mara Santos and Elisa Fernandes for their contribution to this project.

References

[1] S. M. Cardoso. Disfunção Mitocondrial explica doença. *Revista da Associação Portuguesa de Doentes de Parkinson*, vol. 31, 2013.

[2] A. J. Romann, S. Dornelles, N. Maineri, C.R.d.M. Rieder and M.R. Olchik. Cognitive assessment instruments in Parkinson's disease patients undergoing deep brain stimulation. *Dement Neuropsychol*, vol. 6, pp. 2-11, 2012.

[3] [Online] T. N. Tombaugh, "Trail Making Test A and B: Normative data stratified by age and education", 2003. Consulted in May 2015.

[4] [Online] Ammar C. Hamdan, Eli Mara L.R.Hamdan, "Effects of age and education level on the Trail Making Test in a healthy Brazilian sample", 2009.

[5] [Online] G. Veneri, F. Rosini and A. Rufa. Vision and Cerebellum: Evaluating the Influence of Motor Control on Cognitive Execution through Multiscale Wavelet Entropy. *International Journal of Brain and Cognitive Sciences*, 1(2): pages 6-10, 2012.

[6] E. Bigotte, V. Vasconcelos, S. Pires and T. Fonseca. Executive function assessment in Parkinson's disease patients using mobile devices. In proc. *11th Iberian Conference on Information Systems and Technologies*, Las Palmas, Spain, 2016.

Pre-trained ConvNet models as feature extractors and label estimators: A comparative study in large datasets

John Michael
ee11006@fe.up.pt
Professor Luís Teixeira
luisft@fe.up.pt

Faculdade de Engenharia da UP
Universidade do Porto
Porto, Portugal

Abstract

This study explored the viability of out-the-box, pre-trained ConvNet models as a tool to generate features for large-scale classification tasks. A juxtaposition with generative methods for vocabulary generation was drawn (namely probabilistic Latent Semantic Analysis and Sparse Coding) to quantify the performance of these features. Both methods were chosen in an attempt to integrate other datasets (as a form of transfer learning), in the case of the former, and unlabelled data, in the case of the latter. The aforementioned semi-supervised generative models used for the creation of a visual vocabulary were applied to a Spatial Pyramid Matching formulation, to be used with a Support Vector Machine for the image classification task. Lastly, both methods were used together, studying the viability of a pre-trained ConvNet model to estimate category labels of unlabelled images. All experiments pertaining to this study were carried out over two distinct datasets, a two-class set comprised of 25000 images of cats and dogs obtained from Kaggle, which was later expanded into a 5-category dataset with additional fish, whale and galaxy classes (this additional data being again obtained from Kaggle, but also from the ImageNet dataset). The pre-trained models used were obtained from the Caffe Model Zoo, and were trained for the ImageNet dataset challenge.
The comparative study, with multiple trials over each dataset, showed that the pre-trained model was able to achieve the best results for the binary dataset, with an accuracy of 0.945. However, for the extended 5-class dataset, the SPM+SVM method was able to outperform the ConvNet (accuracy of 0.8848 compared to 0.861). Furthermore, when replacing labelled images with unlabelled ones during training, acceptable accuracy scores were obtained (0.87648). Additionally, it was observed that linear kernels perform particularly well when utilized in conjunction with these generative models, allowing for faster training times for the classifiers, while also utilizing less computational resources. This was seen as especially relevant when compared to the ConvNets, which require some days of training even when utilizing 16 Gigabytes of RAM and multiple Nvidia GPUs for computations.

1 Introduction

Image classification is a central problem of Computer Vision and Machine Learning. One of the greatest obstacles faced when handling large volumes of images and video is tied to the fact that, more often than not, the visual data is unlabelled. Whilst unlabelled data is readily available and easy to extract, labelled data is scarce and quite costly to obtain. It's therefore important to find reliable methods which minimize the need for labelled data. A similar argument could be drawn towards using labelled data from categories or applications which are sufficiently similar to the desired one, a form of transfer learning. This can be done indirectly by utilizing pre-trained models as estimators for the class of images from the target dataset. Exploring these various options can help understand how to overcome the scarcity of labelled data.

2 Methodology

The pre-trained ConvNet used to carry out this study (trained on the full Imagenet dataset) follows a slightly modified AlexNet architecture [3], similar to the one proposed for the ZFNet ConvNet [2]. The model was obtained from the Caffe Model Zoo page. The SPM+SVM method [4] utilized a typical image classification pipeline (exemplified in figure 1) and aimed to study the performance of methods for generation of a visual vocabulary, such as Sparse Coding (SC, [5]) and probabilistic Latent Semantic Analysis (pLSA, [1]). Lastly, an attempt to enhance their performance through assistance of the pre-trained ConvNet was studied, by

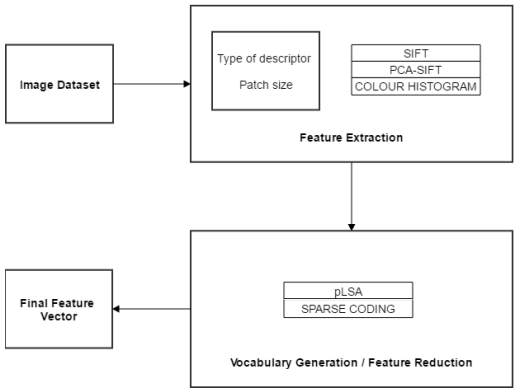


Figure 1: Schematic representing the overall ensemble for image classification

altering the typical SVM cost function:

$$\begin{aligned} &\gamma := \min_{\tilde{\gamma},w} \|w^2\| - C\sum_{i=1}^{n} \varepsilon_i \\ &Subject\ \ to: \\ &\varepsilon_i > 1 \\ &y_i(w^T x_i) \geq 1 - \varepsilon_i \end{aligned} \tag{1}$$

So that the scalar term C becomes a diagonal matrix, where each non-zero entry $c_{(i,i)}$ represents the weight of the ith sample, estimated by class score assigned by the ConvNet. In the training scheme, the sample scores were fixed to $0.5 \times \alpha \times c_{CONV,i} \times \lambda$, where α represents the accuracy score of the pre-trained ConvNet on all the available training labelled data and $c_{CONV,i}$ the class score attributed to unlabelled image i by the ConvNet. This weighting reflects both an estimation of the confidence in the ConvNet's tentative performance (through the parameter α) and an estimation of the tentative classification of the image through $c_{CONV,i}$. After class scores were generated for all unlabelled data, each image was assigned a tentative label through $\max(c_{cat,i}, c_{dog,i}, ...)$, where $c_{cat,i}, c_{dog,i}, ...$ are the normalized scores for the various categories (such as "cat" or "dog"). Values for the average score and maximum score for each class, $c_{max,j}, c_{avg,j}; j = cat, dog$, were also were computed.

2.1 Cross-validation scheme and testing

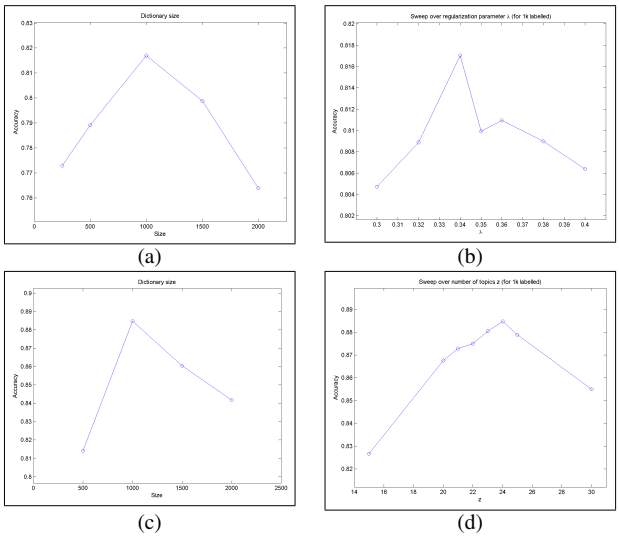
A stratified 4-fold cross validation scheme was used for all the experiments. The final model is obtained by aggregating the results of all the folds through model averaging. Testing each model for performance was done on 12500 images on both the binary and the 5-category datasets. In both cases, the number of images belonging to each category was the same, allowing for isotropic priors which simplify calculations without any meaningful loss of generality. Training and testing was carried out on an Intel i7 4720 processor and 16 GB RAM for the majority of the methods. The exception was training and testing with the pre-trained ConvNet, which required two Nvidia GTX970 GPUs to yield results in a timely manner (training took roughly two days on this hardware).

3 Results

A summary of the main results achieved during the course of this study is presented in table 1. The classical SPM+SVM approached applied to a

Method \ Dataset	Bin	5-class
SPM+SVM	0.82696	0.8848
Pre-trained ConvNet	0.945	0.861
Both methods	0.80928	0.87648

Table 1: Summary table of the best accuracy in each dataset (results presented- SC for the binary case, pLSA for the 5-category case)



(a) (b) (c) (d)

Figure 2: Results of the parameter sweeps, over λ, with $K = 1000$ (a) and over K with $\lambda = 0.34$ (b) for SC in the binary dataset, and over z, with $K = 1000$ (c), and sweep over K, with $z = 23$ (d), for pLSA for the 5-catefory dataset, where K is the dictionary size, λ a constant that controls the sparsity of the vector for SC and z the number of topics for pLSA

visual vocabulary obtained through the Sparse Coding or Latent Semantic Analysis used one thousand labelled images for training. The hybrid ConvNet+generative approach used 500 labelled and 500 unlabelled images. Furthermore, optimization regarding the generative methods used for visual vocabulary generation is present in figure 2, where the effect of overfitting and underfitting can be observed, as expected.

As is customary, ConvNets remain unparalleled in terms of performance for similar image categories. Furthermore, if sufficient computational resources and labelled examples are available, they'll always outperform any other of the presented methods. However, when labelled data is limited, simpler SPM formulations paired up with generative methods offer a viable and computationally lighter solution. These can be used ether in alternative to pre-trained ConvNet models or in conjunction with these, in an ensemble that utilizes the strong points of either approach. One thing to note is that the regularity of classes (that is, how similar elements of each class are) is correlated with accuracy in methods which use SVMs. This can be seen by observing the confusion matrices 2 and 3, noting that the galaxy, fish and whale category have higher regularity. This is an obvious result, as variance within the same class results in more features being captured in each individual image and also in the possibility of some of those features being similar to those present in other classes (as is the case for the cat and dog categories). If a class is very regular, it's also very easily predicted as some very salient, distinct features can be found across all elements of such a class. Considerations about the nature of dataset, the idiosyncrasies of the classification task and limiting factors related to computational resources and time available can weight in favor of some methods and detriment of others. A trade-of was ultimately shown to be present when choosing which method better fits a specific problem.

4 Conclusion

Despite the limitations on available computational resources and the modest time frame in which the study was carried, it was shown that the features extracted by the pre-trained ConvNet model were useful for image classification tasks, yielding comparable accuracy to more traditional methods. Furthermore, when used in conjunction with more typical generative methods for visual vocabulary generation, these allowed to replace labelled data with unlabelled data through the class estimation scheme described without incurring a significant accuracy loss. The results validate the hypothesis that pre-trained ConvNet models can be quite useful in providing an earlier estimation of the class to which an unlabelled image belongs, for posterior use in other models. Furthermore, it was shown that, through some empirical tuning of various weight parameters, the class scores generated by these pre-trained models can offer a satisfactory estimation of the confidence for the tentative labelling provided by the ConvNet. Through the various experiments in both datasets, the performance of multiple SPM models with generative methods creating a visual vocabulary on a dataset with a mixture of labelled and unlabelled data was either kept at a competitive accuracy level. Particularly, the pLSA formulation showed slight performance increases in the labelled and unlabelled mixed sets (of slight over 1%) compared to utilizing only the ConvNet, whilst displaying only very minimal accuracy losses (less than 2% in all cases) compared to the usage of solely labelled data for training. The study showed the benefit of allying the statistical formulation from generative models, which captures richer information in smaller, sparser feature vectors, with out-the-box ConvNet models trained in large, generic datasets. Further, this yielded a decrease in the training time, due to a shorter vocabulary creation step. While ConvNets remained uncontested in performance with sufficient training time and labelled data, these alternative methods successfully utilized unlabelled data for applications where labelled data is scarce and computational resources are limited.

In the future, studying other ways to further integrate these generative methods with ConvNets or explore more elaborate schemes for utilizing the class scores from pre-trained networks might yield even better results.

Labelled \ Predicted	Cat	Dog	Fish	Whale	Galax
Cat	2013	429	-	-	-
Dog	440	2004	-	-	-
Fish	-	-	2249	201	-
Whale	-	-	144	2309	-
Galaxy	-	-	-	-	2481

Table 2: SPM+pLSA confusion matrix for training with 1000 labelled examples for the 5-category dataset, accuracy of 0.8848

Labelled \ Predicted	Cat	Dog	Fish	Whale	Galax
Cat	1998	466	-	-	-
Dog	473	2000	-	-	-
Fish	-	-	2231	202	-
Whale	-	-	189	2261	-
Galaxy	-	-	-	-	2466

Table 3: SPM+pLSA confusion matrix for training with 500 labelled and 500 unlabelled examples for the 5-category dataset, accuracy of 0.87648

References

[1] A. Bosch, A. Zisserman, and X. Munoz. Scene classification using a hybrid generative/ discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:712–727, 2008.

[2] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, Trevor Darrell, and U C Berkeley Eecs. Caffe: Convolutional architecture for fast feature embedding. 2014.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances In Neural Information Processing Systems*, pages 1–9, 2012.

[4] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2:2169–2178, 2006.

[5] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. *Cvpr'09*, 2009.

Single nucleotide variation context in human genome

Vera Enes
vera.enes@ua.pt
Department of Mathematics, University of Aveiro.

João M.O.S. Rodrigues
jmr@ua.pt
IEETA-Institute of Electronic Engineering and Informatics of Aveiro
Department of Electronics, Telecommunications and Informatics, University of Aveiro.

Vera Afreixo
vera@ua.pt
iBiMED-Institute of Biomedicine
IEETA-Institute of Electronic Engineering and Informatics of Aveiro
Department of Mathematics, University of Aveiro.

Abstract

We use the data made available by the 1000 Genomes Project to investigate variation context in the human genome population. We observe that word frequencies in the vicinity of single nucleotide variation (SNV) sites are associated with the type of variation.

1 Introduction

Over a decade ago, the initial sequencing and analysis of what is still considered the reference human genome revealed that less than 2% of the sequence codes for proteins [5]. However, recent evidence suggests that at least 80% of the genome is transcribed or, at least, biochemically active at some point [4]. This evidence highlights the need to understand the biological function of this vast region of the genome, as well as, the evolutionary constraints acting over it. It also highlights the inadequacy of investigating evolutionary constraints by solely considering mammalian conservation criteria, and the need to develop new methodologies to investigate these constraints within a species [13].

The 1000 Genomes Project was a pioneering effort in population sequencing [7]. Their sequence variants with respect to the GRCh37 reference human genome assembly, including single nucleotide variation (SNV), small insertions and deletions (indels), and larger structural variants, are available in the variant call format (VCF, [1, 2, 6]). The single nucleotide variation (SNV) is a genetic variation in a single nucleotide that occurs at a specific position in the genome. There are 6 types of SNVs, which can further be classified as transitions, $C \leftrightarrow T$ and $A \leftrightarrow G$, or as transversions, $A \leftrightarrow C$, $G \leftrightarrow T$, $A \leftrightarrow T$ and $C \leftrightarrow G$.

The genome variations could be a random phenomenon or could be an evolutive/adpative phenomenon. There are some natural questions to ask: Is the variation occurrence position independent from the sequence neighborhood context? Did the neighborhood context presents specific motifs to each type of variation? Are the motifs position in long or short range from the variation position?

Previous work points to a non-random nature of variation occurrences. Variations sites were studied in the mouse genome, and it was concluded that there is a nucleotide bias in the neighborhood of the variation positions, and the association effect decreases with distance from the variation position [11, 14]. In [10], the neighborhoods of 15,110 single nucleotide variations in the bovine genome were analysed, and the authors verified an association between C_pG content and some types of variation. Using the 1000 Genomes Project data, [3] discusses the association between the oligonucleotide neighbourhood variation context (emphasising the heptanucleotide context) and the single nucleotide variations in the human genome.

Here, with the data made available by the 1000 Genomes Project, we present an approach based on the frequency of words in the neighborhood of each annotated variation, to identify new context variation patterns.

2 Material and Methods

We used the GRCh37 reference human genome assembly [9], and version 3 (March 16, 2012) of the Phase 1 integrated variant call set, based on both low coverage and exome whole genome sequencing data from 1,092 individuals [8]. For this study, only the 22 human autosome pairs were considered.

VCF files contain a header followed by variant call records, one per line. Figure 1 shows an extract of the chromosome 1 VCF file from the 1000 Genomes data. (For the sake of legibility, some details were ommited.)

```
##fileformat=VCFv4.1
...
##reference=GRCh37
#CHROM POS   ID  REF ALT  QUAL FILTER INFO FORMAT   HG00096     HG00097
1      10583 ... G   A    100  PASS   ...  GT:DS:GL 0|0:...:... 0|0:...:...
1      10611 ... C   G    100  PASS   ...  GT:DS:GL 0|0:...:... 0|1:...:...
...
1      46402 ... C   CTGT 31   PASS   ...  GT:DS:GL 0|0:...:... 0|0:...:...
```

Figure 1: Excerpt of the chromosome 1 VCF file from the 1000 Genomes phase 1 data. Ommited fields and lines were replaced by ellipsis.

Each record contains several fields of information for a single variation site. The CHROM and POS fields identify the site of variation relative to the reference genome (GRCh37, in this case). The kind of variation is encoded in the REF and ALT fields, which specify the reference allele and alternative allele observed in individual samples. The FORMAT field specifies the encoding used for the remaining fields, each of which contains annotations on a specific individual sample. For example, on the first record, both the HG00096 and the HG00097 samples have `0|0` on the genotype (`GT`) subfield. This means that at this site (position 10583 on chr1), both of these individuals are homozygous with the reference allele, that is, both are $G|G$. On the second record (for position 10611 on chr1), we see that HG00097 is heterozygous `0|1`, meaning it has genotype $C|G$. Other individuals on the same record may be heterozygous `1|0`, meaning $G|C$, homozygous with the reference allele `0|0`, meaning $C|C$, or homozygous with the alternative allele `1|1`, meaning $G|G$. Other fields and subfields in the records include further information, such as the quality or confidence level of the variant calls, but this was not used in this work.

We wrote a short C program, optimized for the specific VCF format used in the 1000 Genomes data, to preprocess the VCF files. This preprocessing consisted of: (1) discarding unwanted fields; (2) selecting only SNV records (rejecting indels and structural variants); and (3) counting samples with each of the `GT` types, `0|0`, `0|1`, `1|0` or `1|1`. This produced much smaller files, with just a few columns, as shown in Figure 2.

```
#CHROM POS   ID  REF ALT C0|0: C0|1: C1|0: C1|1:
1      10583 ... G   A   783   304   0     5
1      10611 ... C   G   1051  37    4     0
1      13302 ... C   T   849   192   45    6
```

Figure 2: Excerpt of the output of the preprocessing stage for the chromosome 1 VCF file. The last four columns show the number of individual samples of each genotype.

The preprocessed output files were then imported into the R software environment [12] for further data normalization and statistical processing. Normalization involved classifying the (REF, ALT) pair into one of the six SNV types, merging the heterozygous counts, and possibly swapping the homozygous counts of each record. The records were then grouped according to the SNV type. Then, the DNA segments in the immediate vicinity to the left and to the right of each SNV site were recovered from the reference genome. An example of the result is shown in Table 1.

Finally, we selected the words of length k located d nucleotides to the right and to the left of each SNV site (see example in Table 1), and produced contingency tables across all of the genome, for each SNV type. This was repeated for words of length $k = 1, 2, 3$, and displacements $d = \pm 1, \pm 2, \ldots, \pm 50$.

CHR	POS	SNV	$A\|A$	$A\|G$	$G\|G$	Left flank	Right flank
1	10583	$A \leftrightarrow G$	5	304	783	*CCCTCGCGGT*	*CT***CT***CCGGGT*
1	54421	$A \leftrightarrow G$	881	202	9	*TAATTGCTTT*	*TC***AC***TCATAT*
1	54490	$A \leftrightarrow G$	13	149	930	*ATACTCTACC*	*GG***CT***TCTGGA*
1	55330	$A \leftrightarrow G$	0	1	1091	*TACTATTTAC*	*CT***TC***AGTAAA*

Table 1: Variation data after preprocessing and normalization. Records are shown only for SNVs of type $A \leftrightarrow G$. The last columns show the left- and right-flanking sequences around the SNV site. Words of length $k = 2$ located $d = +2$ nucleotides to the right of the SNV are highlighted.

The patterns of relative word frequency around the SNV sites are described through the differences to the corresponding average word frequencies in the full data. The association between SNV sites and the type of variation are evaluated with the standard statistical tools: chi-square test and ϕ measure.

3 Results and Discussion

Figure 3 shows nucleotide ($k = 1$) frequency patterns around the SNV sites, grouped by transversions and transitions. The genome sites under study with highest association effect with nucleotide frequencies are in the imediate vicinity ($\pm 1, \ldots, \pm 4$) around the SNV sites. The bias around transitions is much larger than around transversions. Around transversion sites, complementary nucleotides display symmetrical frequency patterns. This is not visible around transition sites.

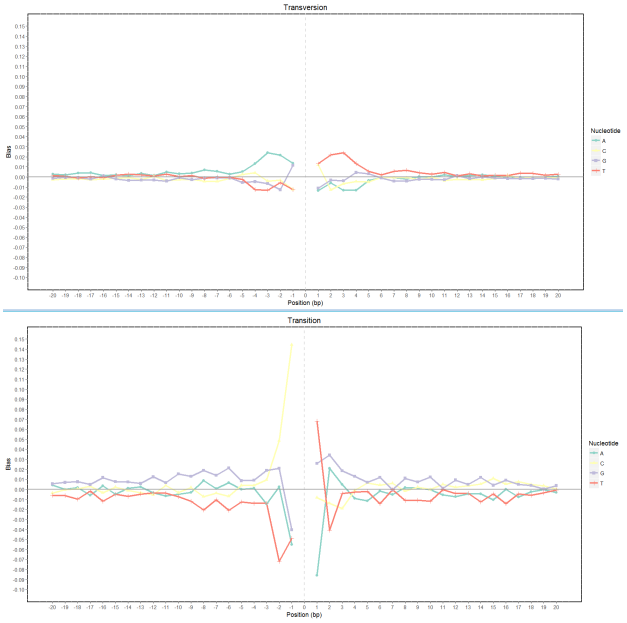


Figure 3: Single nucleotide frequency patterns around transversions (top), and around transitions (bottom).

Each SNV type has specific occurrence context. For example, at the $+1$-site to the right of a $G \leftrightarrow T$ variation, GT is the most favored dinucleotide, and distinct patterns are observed on the left and right flanks (see Fig. 4, top). Around $C \leftrightarrow G$ transversions, the frequency patterns of reverse-complementary dinucleotides seem to be symmetrical (see Fig. 4, bottom).

Funding

This work was supported by Portuguese funds through the iBiMED - Institute of Biomedicine, IEETA - Institute of Electronics and Informatics Engineering of Aveiro and the Portuguese Foundation for Science and Technology ("FCT–Fundaçao para a Ciência e a Tecnologia"), within projects: UID/BIM/04501/2013 and PEst-OE/EEI/UI0127/2014.

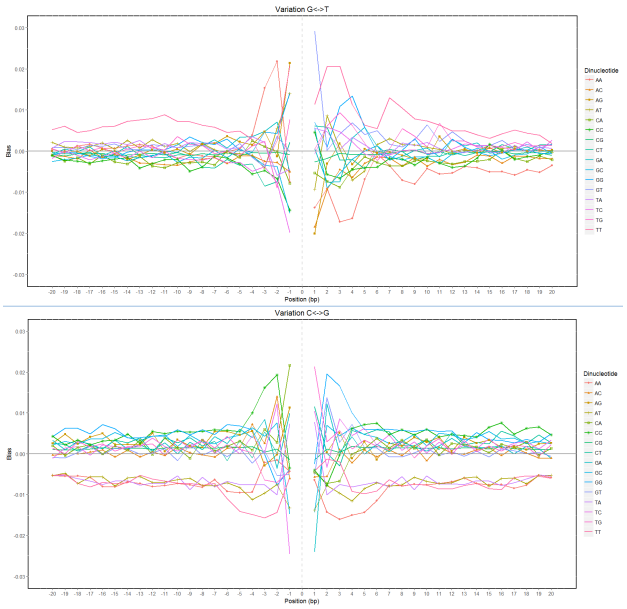


Figure 4: Dinucleotide frequency patterns around $G \leftrightarrow T$ variations (top), and around $C \leftrightarrow G$ variations (bottom).

References

[1] Consortium Project 1000Genomes. An integrated map of genetic variation from 1092 human genomes. *Nature*, 491:56–65, 2012.

[2] The 1000 Genomes Project Consortium (2010). A map of human genome variation from populationscale sequencing. *Nature*, 467: 1061–1073, 2010.

[3] Varun Aggarwala and Benjamin F Voight. An expanded sequence context model broadly explains variability in polymorphism levels across the human genome. *Nature Genetics*, 48, 2016.

[4] The ENCODE Project Consortium. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489:57–74, 2012.

[5] The International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

[6] Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 2011. doi: 10.1093/bioinformatics/btr330.

[7] The 1000 genomes project. 2016. `http://www.1000genomes.org`.

[8] The 1000 genomes project data release: Integrated variant call set for phase 1 version 3. 2016. `ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521`.

[9] GRCh37 Reference human genome assembly. 2016. `ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/ARCHIVE/BUILD.37.3/`.

[10] Zhihua Jiang, Wu Xiao-Lin, Ming Zhang, and et al. The complementary neighborhood patterns and methylation-to-mutation likelihood structures of 15.110 single-nucleotide polymorphisms in the bovine genome. *Genetics*, 180(1):639–647, 2008.

[11] Zackery E. Plyler, Aubrey E. Hill, Christopher W. McAtee, and et al. SNP formation bias in the murine genome provides evidence for parallel evolution. *Genome Biology and Evolution*, 7(9):2506–2519, 2015.

[12] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

[13] L. Ward and M. Kellis. Evidence of abundant purifying selection

in humans for recently acquired regulatory functions. *Science*, 337: 1675–1678, 2012.

[14] F. Zhang and Z. Zhao. The influence of neighboring-nucleotide composition on single nucleotide polymorphisms (SNPs) in the mouse genome and its comparison with human SNPs. *Genomics*, 84(5):785–795, 2004.

Directional Outlyingness applied to distances between Genomic Words

Ana Helena M. P. Tavares[1]
ahtavares@ua.pt
Vera Afreixo[1]
vera@ua.pt
Paula Brito[2]
mpbrito@fep.up.pt
Peter Filzmoser[3]
peter.filzmoser@tuwien.ac.at

[1] Department of Mathematics & iBiMED - Institute of Biomedicine
Universty of Aveiro, PORTUGAL
[2] FEP & LIAAD - INESC TEC
University of Porto, PORTUGAL
[3] Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology, AUSTRIA

Abstract

The detection of outlier curves/images is crucial in many areas, such as environmental, meteorological, medical, or economic contexts. In the functional framework, outlying observations are not only those that contain atypically high or low values, but also curves that present a different shape or pattern from the rest of the curves in the sample. In this short paper, we mention some recent methods for outlier detection in functional data and apply a recently proposed [5] measure, the *directional outlyingness*, and the *functional outlier map* to detect words with outlying distance distribution in the human genome.

1 Introduction

In the functional framework, an outlying observation is not only one that contains atypically high or low values ("magnitude outliers"), but also a curve that presents a different shape or pattern than the rest of the curves of the sample ("shape outliers") [3]. While the first might be easily detected, the latter are often masked among the rest of the curves and thus more difficult to detect.

Different methods for outlier detection in functional data have been developed. Some of those rely on notions of functional depth ([1, 4, 6]. To visualize functional data and investigate the existence of possible outliers, Sun and Genton [6] proposed the functional boxplot and Arribas-Gil and Romo [1] introduced the outliergram. Based on robust principal component scores, Hyndman and Shang [3] proposed graphical tools for visualizing functional data and identifying functional outliers, *e.g.* the bagplot. A very recent approach to detect outlying functions was proposed by Rousseeuw *et al*. [5]. They introduced the directional outlyingness (DO) measure which assigns a robust value of outlyingness to each gridpoint of the function domain, and proposed a procedure that allows detecting outlying functions and outlying parts of a function.

In this work, we consider data arising from the human genome (reference assembly), more precisely, distances between consecutive occurrences of genomic words, and intend to detect words with atypical distance distribution. For fixed word length, the set of 4^k distance distributions can be seen as a sample of curves, which may be treated as functional data. We apply the DO measure to identify atypical distance distributions between genomic words.

1.1 Inter-word distance distribution

Consider the alphabet formed by the four nucleotides $\mathcal{A} = \{A, C, G, T\}$, and let s be a symbolic sequence of length N defined in $\mathcal{A}$. A genomic word, w, is a sequence of length k defined in $\mathcal{A}$. Assuming that the sequence is read through a sliding window of length k, the inter-word distances are the differences between the positions of the first symbol of consecutive occurrences of that word. For example, the inter-CG distances for the DNA sequence $s = ACGTCGATCCGTG$ are 3 and 5.

For each word w, we can define the inter-word distance distribution, f_w, associated with a genomic sequence. In sequences generated by a random process it is expect that distance distributions between genomic words are well fitted by some kind of exponencial law. However, in real genomic sequences we observe distances with peak frequencies and non-expected behaviours.

1.2 Directional Outlyingness

Rousseeuw *et al*. [5] proposed a procedure to detect outlying functions or outlying parts of a function, assigning a robust value of outlyingness to each gridpoint of the function domain. Based on the Stahel-Donoho outlyingness of a point $y \in \mathbb{R}$ relative to a univariate sample $Y = \{y_1, \ldots, y_m\}$, they introduced the notion of *directional outlyingness* (DO), which takes the possible skewness of the distributions into account. Quoting the authors, the main idea is "to split the sample into two halfsamples and then to apply a robust scale estimator to each of them" [5, *pag.3*],

$$DO(y;Y) = \begin{cases} \frac{y - med(Y)}{S_a(Y)} & \text{if } y \geq med(Y) \\ \frac{med(Y) - y}{S_b(Y)} & \text{if } y \leq med(Y) \end{cases}, \tag{1}$$

where S_a and S_b are robust scale estimates for the subsample of points above and for the subsample of points below the median, respectively[1].

The DO of a point $\mathbf{y} \in \mathbb{R}^n$ relative to a n-variate sample $\mathbf{Y} = \{\mathbf{Y_1}, \ldots, \mathbf{Y_1}$ is defined by means of univariate projections, applying the principle that a multivariate point is outlying with respect to a sampleset if it stands out in at least one dimension,

$$DO(\mathbf{y};\mathbf{Y}) = \sup_{v \in \mathbb{R}^n} DO(\mathbf{y}'v;\mathbf{Y}'v)\,, \tag{2}$$

Due to the impossibility of projecting on all directions, the computation of multivariate DO relies on approximate algorithms.

Consider a function x and a functional dataset $X = \{X_1, \ldots, X_m\}$, formed by n-variate functions with univariate domain. At each domain point, t, it is possible to compute the DO of $x(t)$ with respect to the set of values taken by the other functions in the same domain point. Computing a kind of average of those values, a global outlyingness measure of x with respect to X may be achieved. The *functional directional outlyingness* (fDO) of a function x with respect to the functional dataset X, proposed by [5], is defined as

$$fDO(x;X) = \sum_{j=1}^{T} DO(x(t_j);X(t_j))W(t_j)\,, \tag{3}$$

where $W(.)$ is a weight function, which sums one, and $\{t_1, \ldots, t_T\}$ is a discrete set of points of the domain where the functions are observed. The variability of the DO values of a function x is measured by

$$vDO(x;X) = \frac{stdev_j(DO(x(t_j);X(t_j)))}{1 + fDO(x;X)}\,. \tag{4}$$

To visualize the outliers the *functional outlier map* (FOM) is used, a graphical tool firstly proposed in [2] and extended to the DO measure by [5]. The FOM shows a scatter plot of the pairs (fDO, vDO) associated with each curve Y_i, and a fence, drawn from a cutoff rule discussed in [5], which allows puting outliers in evidence. Points in the lower left part of the FOM represent regular functions, holding central positions in the data set. Points in the upper left have low fDO and high vDO, which may be associated with functions with local outliers. Points in the upper right part of the FOM have high fDO and vDO, corresponding to functions which deviate strongly from the majority of the sample.

The method may be applied to multivariate functional data, from univariate curves to images and video data.

[1] The authors used a one-step M-estimator with Huber ρ-function, among many available robust estimators, due to its fast computation and favorable properties [5, *pag.*4].

2 Experimental Results

2.1 Data set

In this study, we used the complete DNA sequences of reference assembly for human genome (GRCh38.p2) downloaded from the website of the National Center for Biotechnology Information. We processed the assembled chromosomes available as separate sequences and studied every word formed by k consecutive nucleotides, with $1 < k \leq 5$.

We computed the inter-word distance distribution of each word, f_w. The dataset contains functions with irregular behaviour revealing several unexpected strong peaks, as the word length increases. The rates of change of the curves may comprise important features on the shape of the data. The inter-word distance distributions were treated as functional data and the dynamic behaviour of the curves was incorporated, by numerically computing their first derivative. To resume, for each word length k, we have a functional dataset formed by 4^k bivariate functions, which response is f_w and its derivative. Since the domains of the curves may be different, we define a cutoff distance, $d^k_{\max}$, associated with each word length.

The computations were performed using the R language. For computing DO and fOM we used R-code provided by Rousseeuw *et al*. at *http://wis.kuleuven.be/stat/robust/software*.

2.2 Detection of outlying inter-word distance distributions

In the present context, the detection of outlier functions obviously depends on the cutoff of the function domain, $d^k_{\max}$. In this first exploratory study, we perform the analysis considering several cutoff distances.

For the dinucleotide case, $k = 2$, the dataset consists of 16 functions defined over a discrete interval (figure 1, top left). We observe that, for short distances, the f_{CG} curve (red) deviates from the other curves. For word length 3, the dataset comprises oscillating functions, but with no evidence of strong peaks. As the word length increases, several distributions have a more expressive oscillating behaviour revealing strong and unexpected peaks. Figure 1 (bottom) shows the f_w for all words of length $k = 5$, where we observe the existence of peaks along a substantial part of the domain.

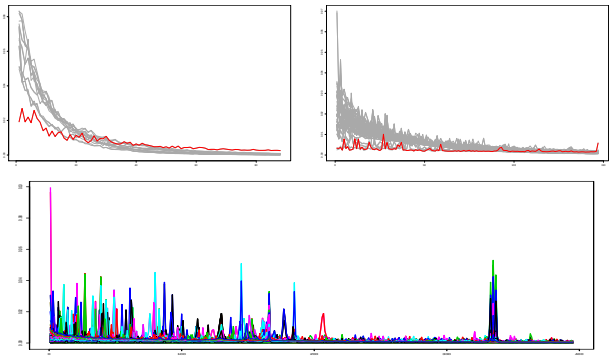


Figure 1: Inter-word distance distributions for different word lengths: $k = 2$, formed by 16 curves, $d^2_{\max} = 90$ (top left); $k = 3$, formed by 64 curves, $d^3_{\max} = 150$ (top right); $k = 5$, formed by 1024 curves, $d^5_{\max} = 400$ (bottom).

The FOMs in figure 2, for the $k = 2$ dataset, show that *CG* data have both high fDO and high vDO. Indeed, for short distances, the *CG* curve deviates from the other curves. However the identification of this curve as outlier depends on the $d^2_{\max}$ value. For $k = 3$, the procedure allows identifying the existence of distributions with both high fDO and high vDO (figure 3, left), which correspond to "flat" distributions, *i.e* distributions with under represented short distances. For $d^3_{\max} = 150$, the TCG curve is identified as outlier (figure 1, middle, in red). Increasing $d^3_{\max}$, other curves with the same behaviour are flagged as outlying functions.

The most interesting case in our analysis is the $k = 5$ dataset. This functional dataset comprises a large proportion of distributions with strong and unexpected peaks, which occur at short and long distances. Furthermore, it reveals clusters of distances where different functions reach unexpected strong peaks. Figure 3 (right) shows the resulting FOM, which reveals the presence of 17 outlying cases, though 10 of them are relatively

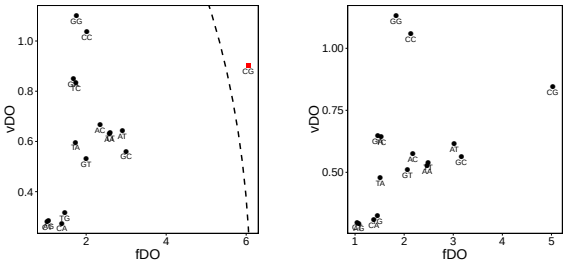


Figure 2: FOM of the $k = 2$ dataset. The detection of outliers depends on the function domain cutoff: for $d^2_{\max} = 90$, one point is flagged as outlier (left); for $d^2_{\max} = 80$ there are no outliers (right).

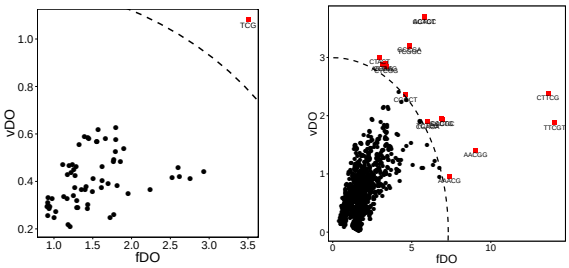


Figure 3: FOM of the inter-word distance distributions data: $k = 3$ data reveal one outlier (left); $k = 5$ data reveal 17 outliers (right).

close to the fence. Analysing the flagged cases one by one, we conclude that the method captures curves with peaks at subdomains where no other peak occurs, as well as curves whose pattern strongly differs from the majority. The two points in the middle right - $CTTCG$ and $TTCGT$ - correspond to functions that deviate strongly from the majority of the curves, they are "shape outliers". Points in the upper left - $AGTGC$, $GCACT$, $GCCGA$, $TCGGC$ - correspond to functions with low fDO but highest vDO values, with outlying behaviour in a small part of the domain. Indeed, the AGTGC curve shows a peak frequency around distance 210. Despite the low peak magnitude, it is located in a interval of the domain with absence of peaks (figure 4, right). Figure 4 (left) confronts the TTCGT curve with the complete set of functions, exposing an unusual curve pattern.

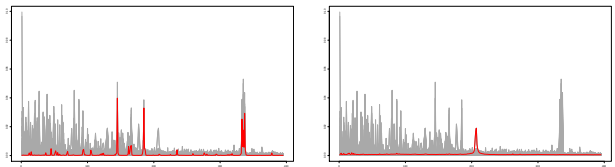


Figure 4: Outlying distributions:; a shape outlier for $w = TTCGT$ (left); local outlier for $w = AGTGC$ (right).

3 Conclusions

The preliminary results indicate that the DO procedure is promising for our problem, puting in evidence outlying inter-word distance distributions masked among the rest of the curves. In the case where the functional dataset comprises a large proportion of functions with strong peaks, spreading over a large part of the domain (*e.g*. $k = 5$ dataset), it is difficult to detect outlying behaviours. The method was able to capture outlying functions distinct from magnitude outliers, highlighting curves whose shape strongly differs from the majority. In particular, it allowed detecting functions with peaks at subdomains where no other peaks occur, as well as functions with several strong peaks. Further analysis will be performed for longer words; future work will investigate the relation between the cutoff in the functions domain and cutoff values for outlier detection.

4 Acknowledgements

This work was supported by Portuguese funds through the iBiMED-Institute of Biomedicine and the Portuguese Foundation for Science and Technology (FCT) within projects UID/BIM/04501/2013 and UID/EEA/50014/2013. AT is supported by FCT PhD fellowship PD/BD/105729/2014. PB is also financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961.

References

[1] Ana Arribas-Gil and Juan Romo. Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4): 603–619, 2014.

[2] Mia Hubert, Peter J. Rousseeuw, and Pieter Segaert. Multivariate functional outlier detection. *Statistical Methods & Applications*, 24 (2):177–246, 2015. (with discussion).

[3] Rob J. Hyndman and Han Lin Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010.

[4] Sara López-Pintado and Juan Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104 (486):718–734, 2009.

[5] Peter J. Rousseeuw, Jakob Raymaekers, and Mia Hubert. A measure of directional outlyingness with applications to image data and video. *arXiv preprint arXiv:1608.05012*, 2016.

[6] Ying Sun and Marc G. Genton. Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334, 2011.

This book of proceedings
for RecPad 2016
has been automatically created by
an advanced intelligent system
based on neural networks,
available at IEETA's laboratories,
and typeset using
LaTeX.