Statistical and Cognitive Spatial Mapping Applied to Face Analysis

Vagner Amaral *, Gilson A. Giraldi † and Carlos E. Thomaz *

* Department of Electrical Engineering, Centro Universitário da FEI
Av. Humberto de Alencar Castelo Branco 3972, São Bernardo do Campo, São Paulo, Brazil
Email: vamaral@fei.edu.br and cet@fei.edu.br

† Department of Computer Science, LNCC
Av. Getúlio Vargas 333, Petrópolis, Rio de Janeiro, Brazil
Email: gilson@lncc.br

***Abstract*—The natural process of pattern recognition has always played a key role in the species survival. For human beings the face recognition has become an essential tool in interpersonal relations and this subject always received attention in the computer vision area due to its applicability. It has been proposed recently, with promising results, a spatial mapping technique that highlights the most discriminating facial features employing some task driven information. The aim of this paper is to compare two distinct methods to define such spatial maps related to some recognition tasks. The first relies on statistical significant differences from the samples and the second on gaze data calculated from eye tracking experiments carried out.**

***Keywords*-task driven; statistical; cognitive; LBP; face analysis;**

I. Introduction

The natural process of pattern recognition has always played a key role in the species survival [1]. For human beings the face recognition process has become an essential tool in interpersonal relations and promoted the development of research in order to understand this phenomenon. In the computer vision and pattern recognition areas this subject always received attention due to its applicability and scope. In the beginning, owing to technological limitations, techniques for geometric facial features extraction were employed [2]. The advancement of computing power enabled the development of holistic approaches using statistical methods [3], [4], [5], [6]. In the last few years a method called Local Binary Pattern (LBP) has been successfully used in face analysis research [7], [8], [9], [10]. Many of these studies have been based on the idea of improving feature extraction via subsequent robust classifiers, ignoring the contribution provided by the contextual information [7], [8], [10]. However, an approach recently proposed [11] has used a spatial mapping technique with promising results, highlighting the most discriminating facial features using some priori information from data mining.

The aim of this paper is to analyse and compare two distinct approaches to define relevant facial features related to some recognition tasks. The first method relies on statistical significant differences from the samples and the second on gaze data calculated from eye tracking experiments carried out. We expect to show as a result that the incorporation of priori information from statistical and cognitive mining into computational algorithms can improve the accuracy of face analysis systems. More specifically, in this study we investigate initially the spatial relevance of facial features in gender and facial expression classifications.

This paper is organized as follows. Next, in section 2, we review the LBP and show the spatial mapping approaches considered in this study. Then, section 3 describes the face image database used to evaluate the effectiveness of the spatial mapping. Perception and computational experiments, and their corresponding results, have been explained in sections 4 and 5, respectively. Finally, in section 6, we conclude the paper, discussing its main contribution and future works.

II. Local Binary Patterns

Initially developed as a texture operator [12], the LBP has been widely employed in face image processing due to its low computational complexity and invariance to monotonic gray level changes [13]. In short, the original approach labels the pixels of an image to encode the local structure around each pixel by thresholding the neighbourhood by its center pixel value. Then, the neighbor values are concatenated as binary number and converted to decimal to label the central pixel.

The output image is divided in R_j regions, $j = 1, 2, ..., N$, usually arranged in a regular grid. The textures descriptors are taken out from each region R_j by their histograms of LBP labels that are grouped in a single feature vector. In the classification process the distinct relevance of physiognomical features are often emphasized [14]. Therefore specific w_j weights are defined for each R_j region. In this work, for example, we used a Chi-Square distance [13]:

$$\chi^2_w(x, y) = \sum_{i,j} w_j \frac{(x_{i,j} - y_{i,j})^2}{x_{i,j} + y_{i,j}}, \tag{1}$$

where x and y are feature vectors to be compared, $x_{i,j}$ is the i histogram bin corresponding to j-th region and w_j its pre-defined weight.

A. Statistical Approach

The possibility to emphasise some physiognomical features allow us to improve the classification step. Thus, a recent work proposed a statistical method that highlights more relevant facial regions in according to the task, employing statistical differences extracted from pixel intensity of samples [11]. This approach consists in to calculate the t-Student test from two distinct face image sample groups, as follows:

$$T = \frac{\overline{X_1} - \overline{X_2}}{S_{X_1X_2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \tag{2}$$

where X_1 and X_2 are face image groups, n_1 is the total number of samples from group X_1 and n_2 is the total number of samples from group X_2. $S_{X_1X_2}$ is given by:

$$S_{X_1X_2} = \sqrt{\frac{(n_1-1)S^2_{X_1} + (n_2-1)S^2_{X_2}}{n_1+n_2-2}}, \tag{3}$$

where $S^2_{X_1}$ and $S^2_{X_2}$ are the variances of the X_1 and X_2 groups, respectively.

In the uniform segmentation procedure the map is divided in a regular grid, composed of rectangular regions. Then, for each j region, we calculate the absolute mean value for T and apply this information as w_j weight in Chi-Square distance (Equation 1) to compare two feature vectors x and y.

B. Cognitive Approach

In order to explore spatial mapping possibilities we have proposed a new approach that employs the human visual strategy for face recognition and classification tasks. Similar to method showed by Ramamurthy [15], this approach provides a spatial gaze map calculated from eye tracking experimental data, that allow to rank the facial regions by their importance, in according to the task performed. These data consist of RGBA images called opacity maps, that represent the most relevant facial regions by the alpha channel intensity. To generate the spatial gaze map we have calculated the mean opacity map from a specific task, i.e. from the facial expression and gender analysis. Figure 1 shows the mean opacity maps overlapped on a face image.

(a) (b)

Fig. 1. Mean opacity maps of human visual strategy highlighting the most relevant face regions in according to the task: a) Face expression; b) Gender.

III. Face Database

To analyse the performance of statistical and cognitive spatial mapping for automatic gender and facial expressions classification we needed a frontal image face database, composed of individuals of both genders and containing variations in neutral and smiling facial expression. Therefore we used a public available sample source that meet the necessary requirements for the experiments called FEI Face Database [16]. We used 200 subjects, 100 women and 100 men. Each subject has two samples, one with neutral facial expression and the other with smiling facial expression, providing a total of 400 frontal face images to perform the gender and facial expression experiments. All the images were pre-processed to reduce the sample variability by the following steps: image rotation until to align the both pupils with the horizontal axe; resize to adjust the interpupillary distance to a pre-defined value; cutting to specified measures; conversion to grey scale, between 0 and 255; and finally histograms equalization of pixel intensities.

IV. Perception Experiments

To acquire gaze data about face recognition tasks and generate spatial gaze maps about gender and expression classification tasks some eye tracking experiments were carried out in this study.

A. Participants

A sample population of 43 volunteers, between 18 and 50 years, from different ethnicities and both sex, with normal vision or corrective lens, consisting of students, teachers and community members of Centro Universitário da FEI, participated in this study. All of them signed a consent form and were instructed about the procedure to be performed during the experiments.

B. Setup

The eye tracker device used in this study is a Tobii TX300 that is supplied with a removable 23” TFT monitor. The configuration, monitoring and data collection were managed by the Tobii Studio on a notebook Core i7, with 16Gb RAM memory, running Windows 7. Eye position data was sampled at 300Hz and 1280x1024 resolution was used in both displays.

C. Stimulus

Face stimuli were randomly chosen from the FEI Face Database to create two subsets with 60 frontal face images each one. Both subsets are composed by 30 males and 30 females, however the first subset used for gender classification tasks has only neutral expression face images and the second subset, used for facial expression classification tasks, has a half of samples for each gender with neutral facial expression and another half with smiling facial expression. To improve the stimuli visualization, in the eye tracker display we resize all the images to 512×512 pixels.

D. Procedure

Subjects were asked to sit facing the eye tracker display near enough to be detected by the infra red sensors, approximately 50cm. After inquire the participant about their comfort we perform the gaze calibration. And we repeat this step until the calibrate status be suitable. Before start the eye tracking experiments the subjects filled a demographic form, about their gender, ethnicity, age and motor predominance, to provide rich data for future works.

The initial screen of both experiments show the instructions, like keyboard functions and stimuli visualization time. After volunteer's command each stimuli is randomly chosen and remains five seconds on the monitor. Then a feedback screen is displayed for undefined time to wait the keyboard response. Before each stimuli a fixation cross appears for 1 second in the middle of screen to help the gaze orientation. At the end of the test a acknowledgement screen is displayed. These steps are illustrated in Figure 2.

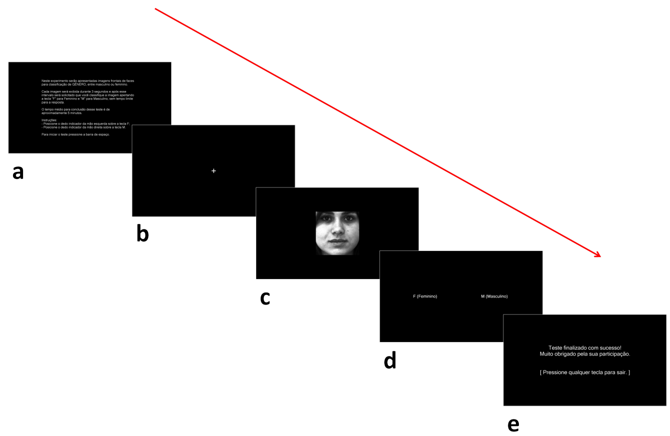

Fig. 2. Eye tracking experimental setup: a) instructions; b) fixation cross; c) stimuli; d) feedback screen; e) acknowledgements

First we presented the gender classification stimuli and then the expression classification stimuli, between these trials we did a little pause to the participants rest. Each test runs until to use the set of samples completely and before each of them we do a gaze calibration again.

E. Data Extraction

In order to provide relevant analytical information we extracted a lot of interaction data using Tobii Studio export tools, including the classification response and mean opacity maps for each stimuli grouped by task.

F. Classification Results

After analysing the extracted eye tracking experimental data we got 97% mean rate for gender classification and 92% mean rate for facial expression classification. In gender analysis task six participants obtained 100% of accuracy and the single worst result obtained 90% of accuracy. In face expression analysis the best result was 98% of accuracy and the single worst result was 80% of accuracy.

V. Computational Experiments

To compare the statistical and cognitive spatial maps in different face analysis tasks we performed an automatic gender and facial expression classification experiments.

A. Setup

In this study all the algorithms were developed and executed on Windows 7 64 bits, using Python 3.4 and Spyder environment, installed by WinPython portable package, containing scientific and mathematical processing libraries.

B. Procedure

We generated the task driven spatial maps for gender and facial expression classification. More specifically, we calculated the t-Student test for the pairs of sample groups that represent particular tasks to obtain the statistical maps and we calculated the mean opacity maps for each task, to obtain the cognitive maps. We normalise the maps range values between 0 and 1, to compare the relevance of physiognomical features in statistical and cognitive approaches. Then we divided them in a 8×8 grid and calculated the mean value for each quadrant to use these values as a spatial weights w_j in Chi-Square distance (expression 1) for histogram comparison. Figure 3 shows all the task driven spatial maps and their corresponding grids.

Then, we arranged the samples in four classification groups. male; female; neutral expression; smiling expression. And performed the classification procedure with the specific task driven maps. Each sample was removed from their group and compared to all another samples to identify the nearest neighbour by Chi-Square distance as classification criteria.

C. Classification Results

In this section, we present the automatic classification accuracy by task driven approaches using statistical and cognitive prior information. These two approaches were verified with four different spatial maps, showed in Figure 3. We included standard LBP accuracy, which ignores the spatial relevance, to highlight the influence of the weighted maps, and human classification rates. The results are shown in Table I.

TABLE I
GENDER AND FACIAL EXPRESSION CLASSIFICATION RATES.

Task	Human Mean Rates	Standard LBP	Task Driven Spatial Maps	
			Statistical	Cognitive
Gender	97%	95%	98%	89%
Expression	92%	54%	83%	69%

VI. Conclusion and Future Works

In this paper we have compared two spatial mapping approaches that emphasize physiognomical features in according to face analysis tasks performed intending to improve the automatic face classification processes using LBP. For this we employed a statistical method to extract relevant information from the data, described in recent works [11], and we proposed a cognitive based method that uses eye tracking experimental data as a task driven information in face analysis algorithms.

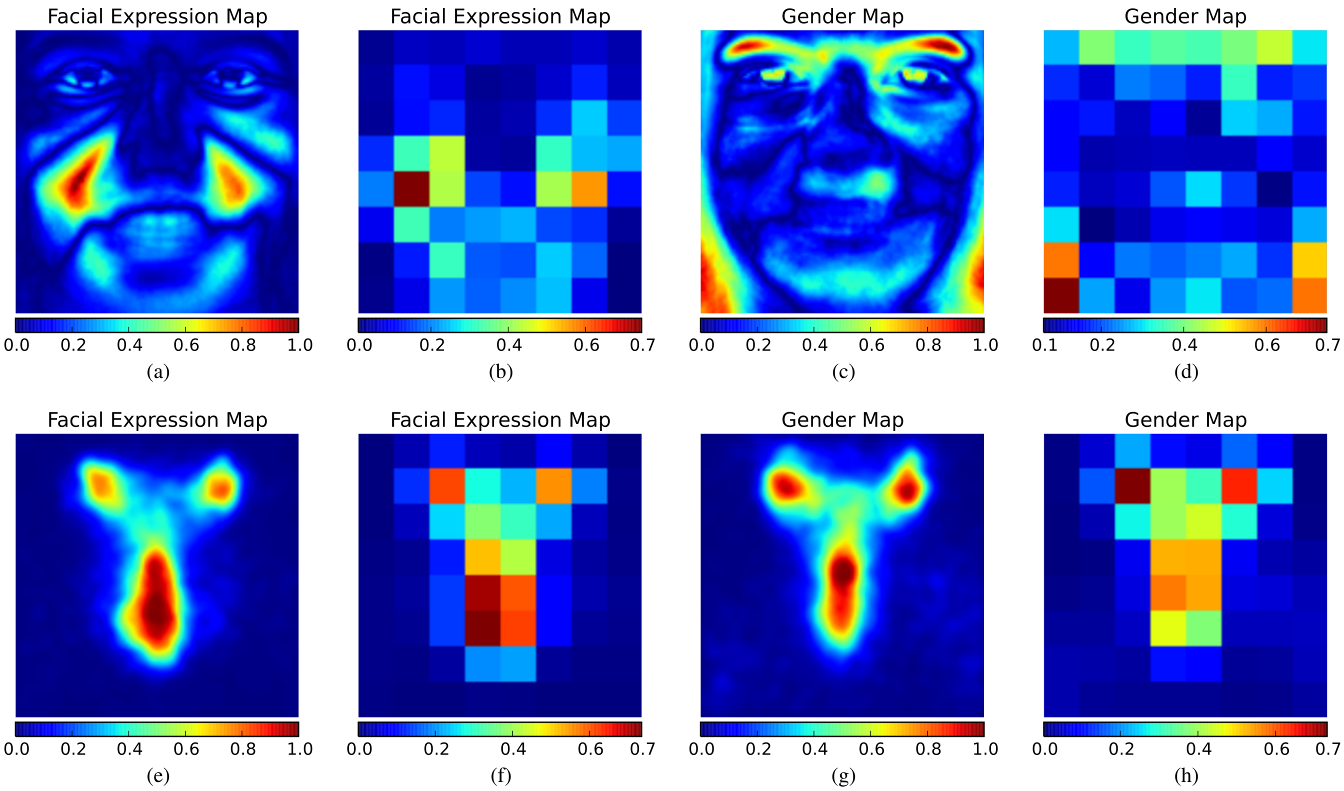

Fig. 3. Task driven spatial maps employed to emphasise some physiognomical features than others. In the first line we show the statistical maps and below the cognitive maps, side by side with their respective grids of w_j spatial weights.

We have verified as an initial contribution that the statistical and cognitive maps, pictured on Figure 3, are completely distinct. The human face recognition accuracies were better or equivalent to computational classification rates. The statistical maps were better than cognitive maps and both of them were better than standard LBP in facial expression task. The statistical gender mapping used a biased data as a discriminant information, i.e. hair in the lower corners of map. However, the facial expression maps used only within face information, as weel as cognitive mapping.

As future works we intend to explore non uniform spatial segmentation methods that allow us to improve histograms relevance, reduce the dimensionality and preserve the configural information of faces. We intend too analyse the demographic data obtained in perception experiments, compare another specific cognitive spatial maps and use masked samples to avoid the lower corner bias.

References

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, 2nd ed. Wiley-Interscience, 2001.

[2] T. Kanade, "Picture processing system by computer complex and recognition of human faces," PhD Thesis, Kyoto University, 1973.

[3] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of Optical Society of America*, vol. 4, no. 3, pp. 519–524, 1987.

[4] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, January 1990.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[6] ——, "Face recognition using eigenfaces," *Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

[7] C. Shan, S. Gong, and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in *ICIP 2005. IEEE International Conference on Image Processing, 2005.* IEEE, September 2005.

[8] ——, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, May 2009.

[9] M. Pietikäinen, G. Zhao, A. Hadid, and T. Ahonen, *Computer Vision Using Local Binary Patterns*, ser. Computational Imaging and Vision. Springer, 2011, no. 40.

[10] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, no. 4, pp. 431–437, March 2012.

[11] V. Amaral, G. A. Giraldi, and C. E. Thomaz, "LBP estatístico aplicado ao reconhecimento de expressões faciais," in *Proceedings of the X Encontro Nacional de Inteligencia Artificial e Computacional, ENIAC'13*, Fortaleza, Ceara, Brasil, October 2013.

[12] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[13] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Analisys and Machine Intelligence*, vol. 28, pp. 2037–2041, 2006.

[14] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[15] B. Ramamurthy, B. Lewis, and A. T. Duchowski, "Eye tracking to enhance facial recognition algorithms," *Conference on Human Factors in Computing Systems 2012*, p. 6, May 2012.

[16] L. L. Oliveira Junior and C. E. Thomaz, "Captura e alinhamento de imagens: um banco de faces brasileiro," Centro Universitário Da FEI, Tech. Rep., June 2006.