LEARNING TO RECOGNIZE GENDER USING EXPERIENCE

M. Castrillón, J. Lorenzo, D. Freire

SIANI
Universidad de Las Palmas de Gran Canaria
SPAIN

O. Déniz

E.T.S. Ingenieros Industriales
Universidad de Castilla-La Mancha
SPAIN

ABSTRACT

Automatic facial analysis abilities are commonly integrated in a system by a previous off-line learning stage. In this paper we argue that a facial analysis system would improve its facial analysis capabilities based on its own experience similarly to the way a biological system, i.e. the human system, does throughout the years. The approach described, focused on gender classification, updates its knowledge according to the classification results. The presented gender experiments suggest that this approach is promising, even when just a short simulation of what for humans would take years of acquisition experience was performed.

***Index Terms*—** Facial analysis, online learning, gender classification

1. INTRODUCTION

A normal human being refines his ability to interpret human faces by being exposed to them during infancy and adolescence [1]. Later his ability to extract information of the human face is impressive. This refinement process has been suggested in experiments of familiar face recognition [2]. Humans are much better at recognizing familiar faces than any other faces. Face familiarity is therefore still a topic of research in the Psychology community, but it seems to be a continuous process acquired by continuous exposure to faces.

The familiarity concept has not been introduced in automatic facial analyzers. Most approaches tackle the usual problems (facial recognition, facial expression classification, gender classification, etc.) neglecting progressive refinement as they are trained offline and rarely updated online. In this sense, standard face recognition tests are designed to recognize a large pool of identities using a single image per individual. This is indeed not the best scenario for humans, we are not so reliable recognizing identities from a single snapshot, e.g. the photo ID [2]. Thus, current automatic systems are trying to recognize faces which are not familiar enough. This situation can be imposed under some circumstances, but it is not the common task in daily interaction. In this context, refining using experience should play an important role.

Incremental learning has been already introduced successfully for systems with autonomous mobility. The humanoid presented in [3] has the ability to learn to recognize people it interacts with. Starting with an empty database, the system follows a completely unsupervised process based on the Eigenface method [4]. In [5] video streams acquired in a single day session were used to extract the information for incremental learning. Another single day experiment based on video streams is presented in [6]. The models are learned online without the need for a pretrained model. In [7] a system based on Incremental Principal Components Analysis (IPCA) was described and tested with a reduced number of identities whose videos were recorded with a difference of one week. A general object learning approach was described [8]. When a new identity is detected by the system, it asks a human supervisor for a label.

This paper tackles the problem of gender classification, one of the most popular automatic facial analysis problems, with high performance using just facial information [9] or combined with other cues [10, 11]. The problem is particularly well suited for our purpose of extensive exposition to the pattern, as the gathering of multiple samples per class is much easier to accomplish than for face recognition. We will simulate the experience by presenting successively faces to the system. Face representation is based on standard techniques, as we are interested not in presenting a new representation approach, but in the model construction procedure.

Representation and classification approaches used are presented in Section 2. Sections 3 and 4 present the experimental results and conclusions respectively.

2. REPRESENTATION AND CLASSIFICATION

2.1. Principal Components Analysis

A classical face representation technique applied to reduce the problem dimensionality is Principal Components Analysis (PCA) decomposition. A normalized image of the target object, i.e. a face, is projected into the PCA space. A number of the resulting coefficients in this space of lower dimension-

Work partially funded by the Spanish Ministry of Science and Innovation and the Castilla-La Mancha Government funds (TIN2008-06068 and PII2I09-0043-3364 respectively).

ality, v_i, is then used to represent the face [4]. The number of coefficients is a configuration parameter of the system.

For an evolving system, it seems to be appropriate to incorporate the new useful information achieved during interaction. For that purpose we have followed the incremental PCA (IPCA) approach described in [12].

2.2. Local Binary Patterns

The Local Binary Pattern (LBP) is an image descriptor commonly used for classification and retrieval. Introduced in [13] for texture classification, they offer invariance to monotonic changes in illumination and low processing cost.

Given a pixel, the LBP operator thresholds the circular neighborhood within a distance by the pixel gray value, g, and labels the center pixel considering the result as a binary pattern. The basic version considers the pixel as the center of a 3×3 window and builds the binary pattern based on its eight neighbors. Rotation invariance is achieved in the LBP based representation considering the local binary pattern as circular.

LBP have already been used to describe facial appearance [14]. Some authors have used histogram based approaches, but other authors have argued that this representation loses relative location information [14, 15]. The application of LBP as preprocessing method, followed by a codification have provided a set of alternatives to the community. Among them, we have selected for our experiments Uniform [13] and Simplified [15] LBP.

2.3. Classification

Support Vector Machines (SVMs) [16] are a set of related supervised learning methods often present in the face analysis literature. They belong to a family of generalized linear classifiers that simultaneously minimizes the empirical classification error and maximizes the geometric margin; hence they are also known as maximum margin classifiers.

In the experiments presented below we compute SVM classifiers based on different input data: 1) PCA coefficients extracted from the gray images, 2) PCA coefficients extracted from the LBP images and 3) histogram values obtained from the LBP images.

When PCA is used (based on the gray or the LBP image) the number of coefficients used to represent the face image in the PCA space is a configuration parameter. Taking a random half of the dataset for training, Figure 1.a shows the results employing a different number of coefficients for a batch classifier. The best performance achieved, around 88%, is within the $40 - 120$ range. In the results presented below, we will report results for 40, 90 and 130 coefficients.

If a histogram representation, based on the LBP image, is employed, the face image is typically divided into non overlapping rectangular blocks, describing the whole face by the concatenation of the histogram collection extracted from each block. We present below the performance using 5×5, 10×10 and 15×15 blocks. Lower and bigger blocks reported poor results.

3. EXPERIMENTS

We have chosen the gender classification problem considering that it is easier to build a large dataset for a two class problem than for a n-class problem, e.g. identity. The main purpose of the experiments is to check if a classifier built using an incremental or evolving approach is comparable in terms of performance with the results achieved with the batch approach.

Representation and classification are performed using the standard techniques described in section 2, but obviously other representation spaces and classifiers can be used. The results have been achieved using a dataset of still images for both the batch and incremental approaches.

To simulate the system evolution, it is successively exposed to a new face, i.e. the system has a *meeting* with a face pattern. Initially the training set of the gender classifier is empty, but it can be enlarged after each meeting. Once the system has training samples, it computes an initial classifier, and will be ready to suggest a label for the next meeting (male/female). A human supervisor verifies the system's proposal, forcing the system to improve the training, using the new pattern, set when it fails. Therefore, the classification results and the image presented will be used by the system to modify, if necessary, the training set and incrementally improve the representation space, using IPCA. Once the training set and representation space are modified, the SVM based classifier is retrained. Training has the largest computational cost, but in our experiments it has always been lower than one second.

After each meeting, the historical performance of the system is computed to confirm if it is improving or not. To avoid the influence of a particular order in the meetings, the database samples were randomly sorted before the experiment and the experiment was performed ten times, reporting in the graphs the averaged results.

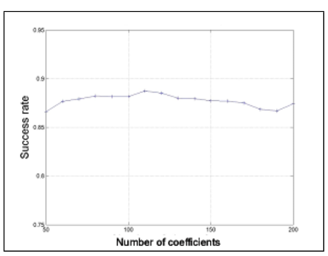
The dataset contains 5847 heterogeneous face images (3380 corresponding to male and 2467 to female) taken from Internet and personal archives. These images have been normalized according to manually annotated eye positions obtaining samples of 59×65 pixels, see Figure 1.b. Finally these images have been labelled according to their gender.

Before presenting the results achieved for online learning, we summarize the performance achieved with batch or offline classifiers. For classification we have used Support Vector Machines (SVMs) that were trained with the following representation approaches:

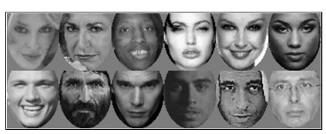
- A number of coefficients (40, 90 or 130) of the PCA space obtained from the original gray images.

- A number of coefficients (40, 90 or 130) of the PCA space computed on the resulting images after preprocessing using LBP. Two different approaches, simplified LBP (SLBP) and uniform LBP (ULBP), have been used.
- A concatenation of histograms (testing different block sizes) based on the resulting uniform or simplified LBP image (ULBPH and SLBPH respectively).

Table 1 presents the averaged results achieved for each configuration after selecting randomly ten times half of the dataset for training and the other half for testing. Different configurations have been used for each representation space, e.g. the number of eigenvectors for the PCA based, and the window dimension for the histogram based approaches were modified in the analysis. Firstly, an ANOVA test with a confidence level of 95% was applied to assess the influence of the representation space in the results considering as factor each representation (PCA, PCA + SLBP, PCA + ULBP, SLBPH and ULBPH) and the different levels of the factor the number of eigenvalues or block size respectively. In this way, it was demonstrated that either the number of eigenvalues or the block size has influenced in the results. For this problem the achieved results suggest that the best approach, achieving roughly 87% (using just a 5% of the dataset for training and 95% for test, its clasification rates was 79%), was the PCA based (making use of 90 coefficients extracted from the gray images), followed by the ULBPH with a 10×10 dimension for each histogram.



(a)



(b)

Fig. 1. (a) Batch classifier gender recognition performance using a different number of coefficients in PCA space. (b) Samples contained in the face dataset.

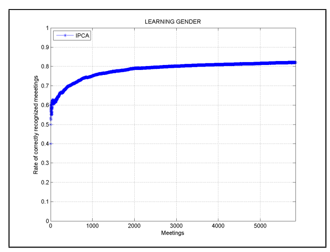
The representation approach getting the best performance is now tested considering the evolving procedure. The averaged results applying the iterative exposition of the system to each single image of the dataset, i.e. roughly 5840 meetings, is presented in Figure 2. The left graph indicates the cumulative rate of correct classification. Given a number of meetings held, it shows the rate of images correctly classified so far. Figure 2.a suggests that after a number of meetings the improvement is less evident. Figure 2.b depicts the number of samples contained in the training set after each meeting. As the training set keeps growing, the graph suggests that more

	PCA			PCA + SLBP			PCA + ULBP		
Eigenvectors	40	90	130	40	90	130	40	90	130
Correct mean	85.8%	87.5%	87.2%	77.8%	79%	78.8%	76.6%	78.7%	78.7%
Correct std	0.31%	0.23%	0.55%	0.73%	0.26%	0.56%	0.74%	0.33%	0.36%

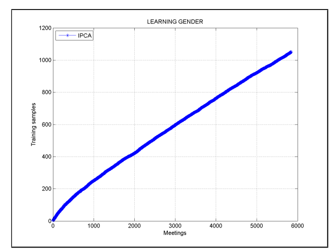
	SLBPH			ULBPH		
Block dimension	5	10	15	5	10	15
Correct mean	57.8%	77.5%	71.9%	57.8%	80.3%	74.8%
Correctstd	0.21%	0.28%	0.29%	0.21%	0.35%	0.73%

Table 1. The dimension of the representation space in the PCA based approaches indicates the number of eigenvectors used for projecting the face image. Related with the histogram based approaches indicates the block size used for each histogram.

meetings are needed to improve the classifier. Figure 3 shows the results processing only those images not contained in the training set, being comparable to the results achieved with the batch or offline classifier but using less than one fifth of the dataset for training.



(a)



(b)

Fig. 2. (a) Gender classification performance evolution. For each meeting, the cumulative rate of correct classification is indicated. (b) Training set size evolution.

4. CONCLUSIONS

We have presented an approach that makes use of its own experience to improve its capability to classify faces. Making use of standard techniques for face representation and classification, an experimental setup has been designed for gender classification. The difficulties of gathering an experimental dataset comparable to years of human visual experience reading faces, have convinced us to address this bi-class problem instead of others as for example face recognition. The gender classification results obtained by the online classifier are similar to those achieved with a batch classifier, while reducing notoriously the size of the training set.

The current stage of the system requires human supervision, in a similar way to a human system during its childhood. However, the evidenced improvement suggests that in a close or medium term, such system could be almost autonomous for this task. Observe that the human supervision used could

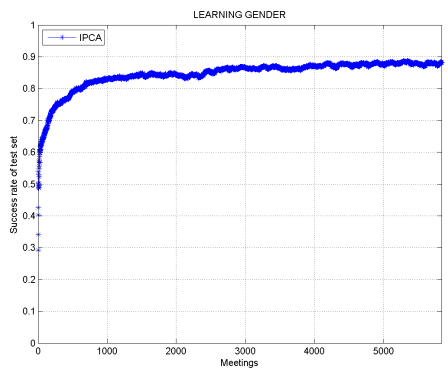


Fig. 3. Gender classification performance evolution testing the complete dataset of static images.

be combined with or substituted by other modalities.

Therefore, we conclude that an automatic system starting with an empty training set presents possibilities of becoming reliable by making use of its own experience. The reader can certainly point out that there are other possibilities to represent and classify faces, but we would like to remind that in this paper we are not focusing on them but on the approach of learning in an evolving way.

Additionally gender classification is only one of the face analysis problems and we are therefore considering to tackling the face recognition problem, after setting up a video stream dataset for multiple identities acquired several times.

5. REFERENCES

[1] Helmut Leder, Gudrun Schwarzer, and Steve Langton, *Development of Face Processing in Early Adolescence*, chapter 5, pp. 69–80, Hogrefe & Huber, 2003.

[2] A. Mike Burton, Rob Jenkins, Peter J.B. Hancock, and David White, "Robust representations for face recognition: The power of averages," *Cognitive Psychology*, vol. 51, no. 3, pp. 256–284, 2005.

[3] Lijin Aryananda, "Recognizing and remembering individuals: Online and unsupervised face recognition for humanoid robot," in *Proc. of IROS*, 2002.

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

[5] Juyang Weng, Colin H. Evans, and Wey-Shiuan Hwang, "An incremental learning method for face recognition under continuous video stream," in *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[6] Liang Liu, Yunhong Wang, and Tieniu Tan, "Online appearance model learning for video-based face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 17-22 June 2007, pp. 1–7.

[7] Seiichi Ozawa, Soon Lee Toh, Shigeo Abe, Shaoning Pang, and Nikola Kasabov, "Incremental learning for online face recognition," in *Proceedings of International Joint Conference on Neural Networks*, Montreal, Canada, July-August 2005.

[8] Heiko Wersing, Stephan Kirstein, Michael Goetting, Holger Brandl, Mark Dunn, Inna Mikhailova, Christian Goerick, Jochen Steil, Helge Ritter, and Edgar Kierner, "Online learning of objects and faces in an integrated biologically motivated architecture," in *International Conference in Vision Systems*, 2007.

[9] Erno Mäkinen and Roope Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 541–547, March 2008.

[10] C. Shan, S. Gong, and P.W. McOwan, "Learning gender from human gaits and faces," in *IEEE Conference on Advanced Video and Signal Based Surveillance*, September 2007, pp. 505–510.

[11] Abdenour Hadid and Matti Pietikäinen, "Combining appearance and motion for face and gender recognition from videos," *Pattern Recognition*, vol. 42, no. 11, pp. 2818–2827, November 2009.

[12] Matej Artac, Matjaz Jogan, and Ales Leonardis, "Incremental PCA for on-line visual learning and recognition," in *Proceedings 16th International Conference on Pattern Recognition*, 2002, pp. 781–784.

[13] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[14] Sébastien Marcel, Yann Rodríguez, and Guillaume Heusch, "On the recent use of local binary patterns for face authentication," *International Journal of Image and Video Preprocessing, Special Issue on Facial Image Processing*, 2007.

[15] Qian Tao and Raymond Veldhuis, "Illumination normalization based on simplified local binary patterns for a face verification system," in *Proc. of the Biometrics Symposium*, 2007, pp. 1–6.

[16] V. Vapnik, *The nature of statistical learning theory*, Springer, New York, 1995.