Identity and Gender Recognition Using the ENCARA Real-Time Face Detector *

M. Castrillón **, O. Déniz, D. Hernández, A. Domínguez

IUSIANI - Universidad de Las Palmas de Gran Canaria - Edificio Central del Parque Científico Tecnológico - Campus de Tafira 35017 LAS PALMAS - SPAIN

Abstract This paper presents recognition results based on a PCA representation and classification with SVMs and temporal coherence.

1 Introduction

This work focuses on real time face analysis. High-level facial analysis processes such as face recognition or gender determination rely on low-level routines that must effectively detect and normalize the faces that appear in the input image in real time.

2 The approach

2.1 ENCARA face detector

The standard face detection problem given an arbitrary image can be defined as: *to determine any face -if any- in the image returning the location and extent of each* [1]. The system used for that purpose must provide real-time data to any recognizer in order to build a complete real-time system.

The real-time face detector used, ENCARA, has been designed attending to: 1) detect frontal faces in video streams, providing fast performance, 2) use only visual information provided by a single camera, 3) employ explicit and implicit knowledge.

The process launches an initial face hypothesis on selected areas in the image. These areas present some kind of evidence that make them valid to assume that hypothesis. Later, the problem is tackled by combining multiple simple classifiers of limited computational cost applied opportunistically in a cascade approach in order to confirm/reject the initial frontal face hypothesis. In the first case, the module results are passed to the following module in the cascade. In the second, the area is rejected. Those techniques are based on contextual knowledge about face geometry, appearance and temporal coherence as detailed in [2].

* The first author is supported by Consejería de Educación, Cultura y Deportes of the Comunidad Autónoma de Canarias, and Beleyma and Unelco through Fundación Canaria Universitaria de Las Palmas. This work was partially funded by research projects PI2000/042 of Gobierno de Canarias and UNI2002/16 of Universidad de Las Palmas de Gran Canaria.

** Email:modesto@dis.ulpgc.es

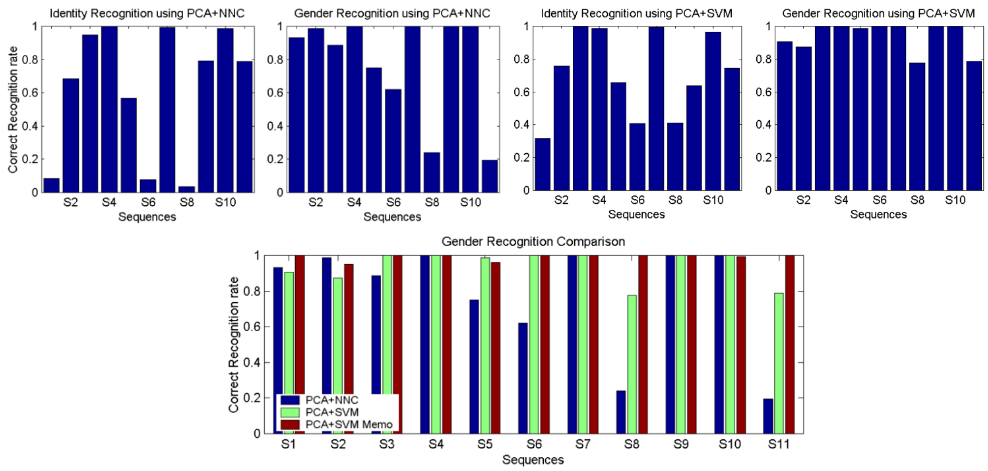

Figure 1. Recognition results. Top left identity and gender recognition results using PCA+NNC, top right using PCA+SVM. Bottom gender results comparison.

2.2 Basic representation/classification techniques

Any facial pattern extracted automatically from a sequence to represent the individual appearance has certainly a high dimensionality. A first option could be to work with the image as a vector where each pixel represents a dimension, but there is high redundancy in this coding. Principal Component Analysis (PCA) (first applied to faces in [3]) chooses the dimension reduction that maximizes the scatter of the projected samples in this space.

Once a representation in a reduced dimension space is obtained, e.g. with PCA, there are different options for classification. Classification can be performed with a simple scheme such as a nearest neighbor classifier (NNC), training a neural network with those representations, etc. Recently, Support Vector Machines (SVMs) have been applied to the face recognition problem.

SVMs [4] allow to represent classes as a set of points in a high dimensional space where the class boundaries can be expressed as hyperplanes. For the linearly separable case, SVMs provide an optimal hyperplane that separates training patterns while minimizing the expected error of novel patterns. For non linear problems, training patterns are mapped into a high-dimensional space using a kernel function. The problem is expected to have a linear solution in that space.

3 Recognition Experiments

Two different datasets have been collected for the experiments. The first set contains different video streams acquired and recorded using a standard webcam. These sequences, labelled S1-S11, of 7 individuals were acquired on different days without special illumination restrictions. The sequences were taken at 15 Hz. during 30 seconds, i.e., each sequence contains 450 frames of 320×240 pixels.

The identity training set is built using random images for each subject. These images are extracted automatically by ENCARA from sequences S2, S3, S4, S5, S7, S10 and S11. Two random patterns are selected from the first ten detections.

The standard PCA representation approach [3] is used for representation. As a first approach the nearest neighbor classifier (NNC) is selected for classification in that reduced space. In the top left most bar graph of Figure 1, identity recognition results for the eleven sequences are presented. For those sequences used to build the training set, the results are considerable better. It must be noticed that sequences of the same subject were not taken the same day. The low rate for sequences not used for training is mainly due to the different lighting conditions used for acquiring the sequences for the same individual.

Gender classification was also applied similarly to [5]. In the second top left most bar graph of Figure 1, the results are presented. It is notorious the failure for sequence S11, used to extract training samples, but in general the approach presents better performance than the identity recognizer for sequences not used to extract training frames except S11.

These results seem promising but it is evidenced that PCA presents problems with illumination changes. Recent developments use more local representations, as for example Independent Component Analysis (ICA) [6]. Recently, the work described in [7] showed that the use of any of both representation spaces (PCA or ICA), and a powerful classification algorithm instead of NNC, such as SVMs, reports similar recognition rates. This work concluded that the classification criteria selection may be more important than the representation used. Therefore, new experiments have been carried out using SVMs as classifier.

Results as presented for identity and gender top right most bars in Figure 1, being in general better even for those sequences not used to extract training samples. The gender classifier using PCA+SVM performs always over 0.75 while the identity classifier performs for sequences used for training over 0.7 while only over 0.3 for sequences acquired with different conditions. In any case the overall performance of this approach is better, in this experiment, using PCA+SVM.

These experiments concern to video streams. Thus, it is not common to suddenly experience an identity or gender change. To confirm the identity of a person, temporal coherence on the results associated with a blob is needed. Assuming temporal coherence for recognition will lead us to avoid instant changes of identity. According to the empirical work in [8], the most robust temporal fusion scheme is majority vote. Using this idea of memory for gender identification, the comparison with previous approaches is presented bottom graph in Figure 1, being the approach that makes use of temporal coherence clearly better, performing over 0.93 for all the sequences.

The second dataset used is larger and contains 48 sequences. Observing the good results achieved for gender classification, a learning set was built with a single pattern extracted from 28 of those 48 sequences (i.e. 20 individuals are not contained in the database). This set is increased using mug shots of approximately 1000 individuals not contained in the sequences. This fact increased the learning set, getting finally 798 samples for male appearance and 231 for female appearance. The real-time face detector analyzed 17529 images from the sequences detecting frontal faces in 9878 of them. Among these detections, 8123 of them correspond to male faces and 1755 to female faces.

For each image gender recognition is performed representing the normalized face in PCA, applying three different schemas to classify: 1) NNC, 2) SVMs and 3) SVMs with memory. Each approach reported respectively an average success rate of 84.63% (PCA+NNC), 91.88% (PCA+SVM) and 98.57% (PCA+SVM with memory).

4 Conclusions and Future Work

The integration of the recognition stage for a real-time classifier does not reduce the latency of the system, typically 20 msecs using a PIV 2.2GHz.

Recognition experiments provide promising results for real-time recognition in desktop scenarios using a reduced set of individuals. These results are very good, 98.57%, for a large database for gender classification integrating SVM and memory for classification.

Future work must tackle the analysis of greater databases and the improvement of individual modelling to check the application of the approach for identity recognition to larger databases.

References

1. Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
2. Modesto Castrillón Santana. *On Real-Time Face Detection in Video Streams. An Opportunistic Approach.* PhD thesis, Universidad de Las Palmas de Gran Canaria, March 2003.
3. M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.
4. V. Vapnik. *The nature of statistical learning theory.* Springer, New York, 1995.
5. Baback Moghaddam and Ming-Hsuan Yang. Learning gender with support faces. In *Proceedings of the fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000)*, May 2002.
6. M.S. Bartlett and T.J. Sejnowski. Independent component of face images: a representation for face recognition. In *Procs. of the Annual Joint Symposium on Neural Computation,Pasadena, CA*, May 1997.
7. O. Déniz, M. Castrillón, and M. Hernández. Face recognition using independent component analysis and support vector machines. In *Procs. of the Third International Conference on Audio- and Video-Based Person Authentication, Halmstad, Sweden. Lecture Notes in Computer Science 2091*, pages 59–64, 2001.
8. O. Déniz, J. Lorenzo, M. Castrillón, and M. Hernández. Estudio experimental sobre la combinación temporal de resultados en el reconocimiento de caras con secuencias de video. In *IX Conferencia de la Asociación Española para la Inteligencia Artificial, Gijón*, pages 59–64, November 2001.