FACE RECOGNITION OF VIDEO SEQUENCES IN A MPEG-7 CONTEXT USING A GLOBAL EIGEN APPROACH

L. Lorente   *L. Torres*

Polytechnic University of Catalonia
Barcelona, Spain
alorente@gps.tsc.upc.es   luis@gps.tsc.upc.es

ABSTRACT

An integral scheme that provides a global eigen approach to the problem of face recognition of still images has been presented in [1]. The scheme is based on the representation of the face images using the so called eigenfaces [2], generated performing a PCA (Principal Components Analysis). The data base used was designed for still image recognition and the corresponding images were very controlled. That is, the test images had controlled expression, orientation and lighting variations. Preliminary results were shown using only a frontal view image by person in the training set. In this paper, we present our first results for face recognition of video sequences. To that end, we have modified our original scheme in such a way that is able to cope with the different face conditions present in a video sequence. The main and final objective is to develop a tool to be used in the MPEG-7 standardization effort to help video indexing activities. The system is not yet fully automatic, but an automatic facial point location is under development. Good results have been obtained using the video test sequences used in the MPEG-7 evaluation group.

1. INTRODUCTION

Face recognition has been object of much interest in the last years [3] [4]. It has many applications in a variety of fields such as identification for law enforcement, authentication for banking and security system access, and personal identification among others. In addition to all these applications, there is an increasing interest to specify standardized descriptions of various types of multimedia information. This description will be associated with the content itself, to allow fast and efficient searching for material that is of interest to the user. This effort is being conducted within the activities of the new standard MPEG-7 (Multimedia Content Description Interface) [5]. It is in this context that face recognition acquires a renovated interest and there is a need to develop new tools that may help the user that browse a data base to answer the following type of query:

Is there any face in this video sequence that matches that of Marlon Brando? The automatic answer to this question is at this time very difficult, and it needs, at least, three stages: segmentation of the sequence in different objects, location of objects that correspond to human faces and recognition of the face. Our research group has been already involved in developing tool for the segmentation [10], [11] and location stages [12] and now we are approaching the face recognition stage. Our final objective is to have a system that automatically segments, locates and recognizes a human face in a video sequence. It has to be emphasized that almost all efforts in face recognition, have been devoted to recognize still images. A very few works have presented results on video sequences [6].

Face recognition of video sequences has many problems as, in general, the person’s face is exposed to very different illumination conditions, different size scales, different face expressions, and specially in many occasions significant parts of the face are occluded and only limited face information is available. In addition, in the MPEG-7 standardization activities effort, the accepted test sequences are in MPEG-1 format what poses additional problems, mainly the small size of the faces and the noise introduced by the MPEG compression algorithm. There is a need then, to develop efficient face recognition schemes which may take into account the CIF format and the low quality present in MPEG-1 sequences. We present in the following some proposals to the problem of face recognition of video sequences.

2. PRINCIPAL COMPONENT ANALYSIS

This work has been supported by Grant TIC98-0422 *of the Spanish Government and by the Hypermedia ACTS European project*

Among the best possible known approaches for face recognition, Principal Component Analysis (PCA) has been object of much effort [2], [4]. In PCA, the recognition system is based on the representation of the face images using the so called *eigenfaces*. In the eigenface representation, every training image is considered a vector $\overrightarrow{X_i}$ of pixel gray values (i.e. the training images are rearranged using row ordering). The full image space is not an optimal space for face description, because all the face images are very similar, and therefore all the face vectors are located in a very narrow cluster in the image space. The goal of the eigenface representation is to find the basis that best represent the image faces. A solution is to apply a PCA analysis over the training faces. The main idea is to obtain a set of orthogonal vectors that optimally represent the distribution of the data in the RMS sense. Using these vectors, a good representation of the faces may be obtained. If the average face is defined as

$$\vec{m} \approx \frac{1}{N}\sum_{i=1}^{N}\overrightarrow{X_i} \qquad (1)$$

these vectors are the eigenvectors of the covariance matrix

$$\underline{\underline{\Sigma_X}} \approx \frac{1}{N}\sum_{i=1}^{N}(\overrightarrow{X_i}-\vec{m})(\overrightarrow{X_i}-\vec{m})^T \; . \qquad (2)$$

Any training image can be obtained as a linear combination from the eigenvectors. These eigenvectors are usually referred to as eigenfaces because they look like faces. Furthermore, good representations of the face images can be obtained using only a low number of eigenfaces.

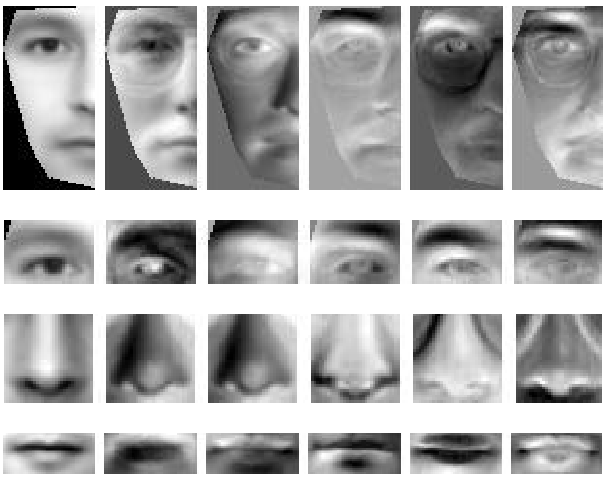

Figure 1. Mean and first eigenfeatures of some parts of the face (left eigensides, left eigeneyes, eigennoses and eigenmouths), using a training set of 92 frontal face images.

PCA is considered one of the techniques that provides the best performance [4]. The main idea of the PCA is to obtain a set of orthogonal vectors (eigenfaces) that optimally represent the distribution of the data in the RMS sense. Once the corresponding eigenfaces are computed, they are used to represent the test and training faces to be identified. The test face (the one to be recognized) is matched to the training face whose eigen representation is the most similar [4].

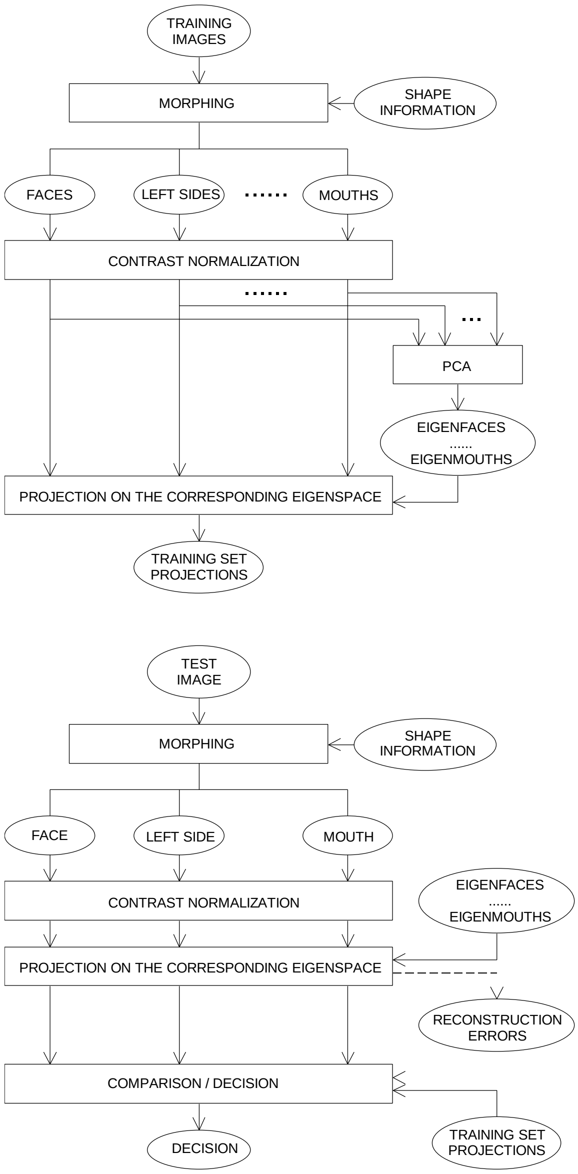

Figure 2. General eigen approach for face recognition: Above: training stage; below: test stage

As in [7], we have extended the eigenface concept to certain parts of the face: both eyes (left and right eigeneyes), the nose (eigennoses) and the mouth (eigenmouth). We have also introduced the new concept of eigenside (left and right), which are eigenfaces generated from the left and right sides of the face (see Figure 1 for some examples). This way, it is possible to avoid certain limitations of the eigenfaces, mainly when some parts of the faces are occluded, or when some conditions such as lateral lighting or facial expression change along the face. The key issue is to take advantage of the fact that the distortions caused by these difficult conditions in the video sequence affect more to some parts of the face than others. Our scheme also includes a morphing stage to minimize the variations in orientation, facial expression and lighting. The morphing stage is crucial in our scheme and has been greatly improved to be adapted to video sequences. The general face recognition scheme is presented in Figure 2.

3. MORPHING STAGE AND IMAGE MODEL

The morphing process gives an image face in which the facial feature points have been moved to standard predefined positions. This is equivalent to a normalization of the shape of the faces. These standard positions correspond to a frontal view with neutral expression and a given fixed size. The morphing technique is based on texture mapping, a widely known technique in computer graphics [8]. The image of a face is split into triangular polygons (based on the Candide model [9]) whose vertices are characteristic facial points. The texture of each original triangle is mapped to the triangle defined by the standard points. The points used in this synthesis stage are obtained from the Candide model that has been previously matched to the face. Although some shape information is lost during this morphing stage, the recognition system greatly improves because the resulting images are normalized in size, expression, and above all in orientation. The morphing is applied to all the faces to be identified, independently of their degree of variation in expression or orientation.

Our original face recognition system uses a facial model with 44 points and 78 polygons (triangles), based on the Candide model. The location of the facial points was done manually. The use of the face recognition system with MPEG-1 video sequences implies that the system must be able to work with small face images. In addition, the final face recognition system must be fully automatic or semi-automatic, a difficult task with small face images with a significant noise level. For all these reasons we have developed a simplified image face model that only needs to know the location of four main facial points (center of both eyes, nose and mouth). The automatic location of these points is an affordable task even with small images. At the present stage of the development the points are located manually. An automatic location algorithm is being developed with the eigenfeature technology.

The simplified model is built from the four main facial points. A model based solely on these four points cannot fit the shape of the face. To correct this, thirteen additional facial points are automatically estimated, so

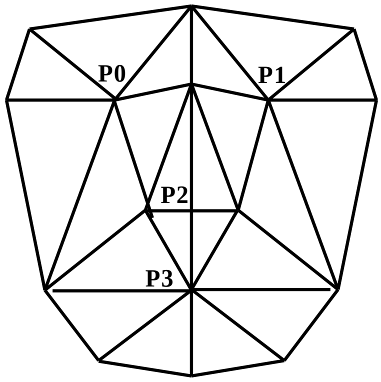

Figure 3. Simplified image face model.

the model has 17 points. The facial points are connected by 22 triangles. All the additional points except 3 of them are located on the edge of the face The three additional inner points are needed due to the great changes of the nose appearance when there are orientation variations. Every additional point is located with the mean of its geometrical relations with the four main points. For every additional point there is also an additional correction to cope with the orientation variations. Figure 3 shows the simplified image face model.

Figure 4. Examples of normalization using the simplified model. With small face images the normalized images have bad quality

Finally, the contrast of the images is normalized, adjusting the pixel values in order to occupy the range [0, 255]. Figure 4 shows some examples of normalization of face images, extracted from the official MPEG-7 test sequences (compressed in MPEG-1 format).

4. RECOGNITION STAGE

In the recognition stage, the corresponding overall eigenfaces, right and left eigensides, right and left eigeneyes, eigennoses and eigenmouths are found for the training and test faces. The normalized Mahalanobis distance is applied between every eigenfeature of the test two and training faces. The Mahalanobis distance between vectors is defined as

$$d\left(\vec{x}, \vec{y}\right) = \sqrt{\sum_{i=1}^{N} \frac{(y_i - x_i)^2}{l_i}}\ , \qquad (3)$$

where y_i and x_i are the components of the vectors and λ_i are the eigenvalues associated to each eigenface (obtained with the PCA). This distance treats variations along all axes (eigenfaces) as equally significant by giving more weight to components corresponding to the smallest eigenvalues. This reasoning seems appropiate since our aim is discrimination. The experiments indicate that this distance is more effective than the Euclidean distance.The global distance between the test and the training image is the weighted sum of the different distances in such a way that the contribution of each eigenfeature to the recognition stage is the same. Weightings factors may be altered to give more importance to certain features than others. The system can also use some thresholds (of reconstruction errors) to decide if a part of the face is useful for the recognition. This should be useful in the cases in which the faces are partially hidden by an obstacle.

5. RESULTS

Tests using the four point model have been conducted with the MPEG-7 test sequences. In particular the sequences news12, news11, music, movie, drama and contest have been used. The images have been separated in two sets: the training set and the test set. The training set contains 59 different images corresponding to 59 different persons and the test set 25 persons. The training images are the images known by the system, so the person ideally must recognize all the images of every face that has an image in the training set. The test set are other images of people represented in the training set. Note that the difficulty of the recognition is due to the number of training images. The training and test image of the persons have been always taken from differents shots of the sequences, although in a few cases some conditions are similar in both shots. Please notice that only a frontal view of the person has been used in the training set. Results are provided in Table 1.

It seems that every eigenfeature gives the best results with a different normalization process. If these results are consistent with larger test sets a different normalization process for every feature of the face might be done. The recognition results are far from the results obtained with controlled databases [1]. This is due to several reasons:

- The rotation and lightning conditions of the training images are neither homogeneus nor controlled.
- The size of many images is too small and there is noise due to the MPEG-1 compression. This problem is more critical in the training images.
- The high rotation angles of the faces
- The variations of rotation and lightning conditions come together in many images of the test and training set
- The presence of obstacles

The weights assigned for the partial distances of every feature in the test are all one. The fine tuning of these weights can lead to a slight improvement.

EIGENFEATURE	RESULTS
GLOBAL	19/25 (76.00%)
EIGENFACE	18/25 (72.00%)
LEFT EIGENSIDE	14/25 (56.00%)
RIGHT EIGENSIDE	15/25 (60.00%)
LEFT EYE	11/25 (44.00%)
RIGHT EYE	14/25 (56.00%)
NOSE	13/25 (52.00%)
MOUTH	4/25 (16.00%)

Table 1. Results of the face recognition system with training (59 images) and test (25 images) sets taken from the MPEG-7 test sequences.

Although the results are still in a preliminary stage, they show that the taken approach will be helpful for the video indexing application. Figure 5 provides some examples of correct and incorrect matches. The left incorrect match is due to two of the problems commented above. The test image is too small, while the training image has lateral illumination (like the test one) that fools the system. These kind of errors can be avoided if the conditions of the training set are controlled.

REFERENCES

[1] L. Lorente, L. Torres, "*A global eigen approach for face recognition*", International Workshop on Very Low

Bit-rate Video Coding, Urbana, Illinois, October 8-9, 1998.

[2] M. A. Turk, A. P. Pentland, "*Face recognition using eigenfaces*", Proceedings of the IEEE Computer Society Conf. on Computer Vision and Patter Recognition, pp. 586-591, Maui, Hawaii 1991.
[3] R. Chellappa, C. L. Wilson, S. Sirohey, *Human and machine recognition of faces: a survey*, Proceedings of the IEEE, Volume 83, No. 5, pp. 705-740, May 1995.

[4] J. Zhang, Y. Yan, M. Lades, "*Face recognition: eigenface, elastic matching, and neural nets*", Proceedings of the IEEE, Vol. 85, No. 9, pp. 1423-1435, September 1997.

[5] ISO/IEC JTC1/SC29/WG11. MPEG Requirements Group. "*MPEG-7: Context and Objectives*", Doc. ISO/MPEG N2460, October 1998 / Atlantic City, USA.

[6] S. McKenna, S. Gong, Y. Raja, "*Face recognition in dynamic scenes*", British machine vison conference, 1997.

[7] A. Pentland, B. Moghaddam, T. Starner, "*View-based and modular eigenspaces for face recognition*", MIT Media Laboratory Perceptual Computing Section, Technical Report 245, 1994.

[8] D. Rowland, D. Perret, D. Burt, K. Lee and S. Akamatsu, "*Transforming facial images in 2 and 3-D*", Imagina 97 Conferences - Actes / Proceedings, Feb. 1997, pp. 159-175.

[9] "*The Candide software package*", Image Coding Group, Linköping University, 1994.

[10] L. Torres, D. García, A. Mates, " A Robust Motion Estimation and Segmentation Approach to Represent Moving Images with Layers", International Conference on Acoustics, Speech and Signal Processing, Munich, Germany, April 21-24, 1997.

[11] F. Marqués, M. Pardàs, P. Salembier, "Codingº1 oriented segmentation of video sequences", in L. Torres, M. Kunt. Editors. Video Coding: The Second Generation Approach, Kluwer Academic Publishers, January 1996.

[12] V. Vilaplana, F. Marqués, "Face segmentation using connected operators", Mathematical Morphology and its Applications to Image and Signal Processing, Amsterdan, pp. 207-214, June 1998.

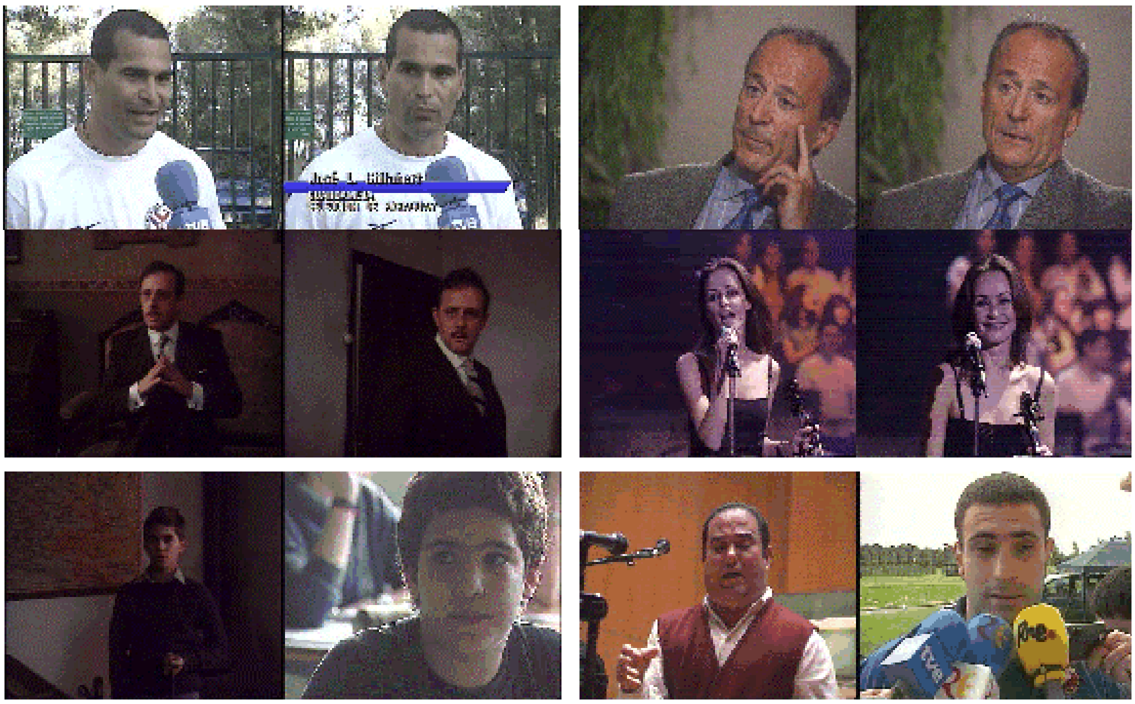

Figure 5. Examples of correct (two above rows) and incorrect (inferior row) matches. For every couple of images, the left image belongs to the test set and the right one to the training set