AUTOMATIC FACE RECOGNITION OF VIDEO SEQUENCES USING SELF-EIGENFACES

L. Torres L. Lorente Josep Vila

Department of Signal Theory and Communications
Polytechnic University of Catalonia
Barcelona, Spain

{luis,alorente,aljosep}@gps.tsc.upc.es

Abstract

The objective of this work is to provide an efficient face recognition scheme useful for video indexing applications. In particular we are addressing the following problem: given a set of known images and given a video sequence to be indexed, find where the corresponding persons appear in the sequence. *Conventional* face detection schemes are not well suited for this application and alternate and more efficient schemes have to be developed. In this paper we have modified our original generic eigenface-based recognition scheme presented in [1] by introducing the concept of self-eigenfaces. The resulting scheme is very efficient to find specific face images and 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. 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 [2] [3]. 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) [4]. 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. 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] [1]. A combined face detection and recognition generic scheme has already been presented in [7].

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

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. 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 a proposal to the problem of face recognition in video sequences based in the concept of self-eigenfaces. Section 2 will present the proposed approach and section 3 will show the results obtained for video sequences.

2. Principal component analysis and self-eigenfaces

Among the best possible known approaches for face recognition, Principal Component Analysis (PCA) has been object of much effort [3], [5]. In PCA, the recognition system is based on the representation of the face images using the so called *eigenfaces.* PCA is considered one of the techniques that provides the best performance [3]. 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 [5].

In a usual eigenface-based scheme for face recognition, such as identification for law enforcement or personal identification, the PCA is performed on a mixture of different training images similar to the unknown images which are to be recognized. However, this approach may be greatly improved when the objective is to find specific person faces within a video sequence. In this case it is more useful to perform a PCA analysis on a set of different views of the same image which is to be recognized. We have called this technique a self-eigenface approach. Let us clarify it with a very simple example. Assume that a video sequence is to be indexed and we want to find out whether or not three specific persons are in the sequence. The training images consist of different views of the same persons. Then a different PCA is performed on each set of views of the same person giving three different sets of eigenfaces, one for each of the persons who want to be recognized. It can be noticed that the main difference with the normal eigenface approach [5] is the number of different sets of eigenfaces.

The test stage has to be modified accordingly. In the self-eigenface approach, each image to be recognized is projected and reconstructed using each one of the sets of the different eigenfaces. In the example above, three different reconstructions are found, one for each set of eigenfaces. The unknown image will be said to match a particular one, when the corresponding reconstruction error using a set of eigenfaces be minimum. Figure 1 shows a simplified block diagram of the proposed approach.

An important step in the process is the normalization approach used to minimize the differences due to changes in size, expression and orientation of the training and the test image set. In our previous proposal [1], a manual normalization using the Candide image model [8] was done. However, although this technique provides very good results, it is very time consuming which is unacceptable in video indexing applications. Therefore we have chosen to sacrify performance but to gain in speed by using a very simple technique. The normalization stage defines one standard

height and width and the images are resampled to fit these measures using a bilineal interpolation. A histogram normalization is also applied to the original images.

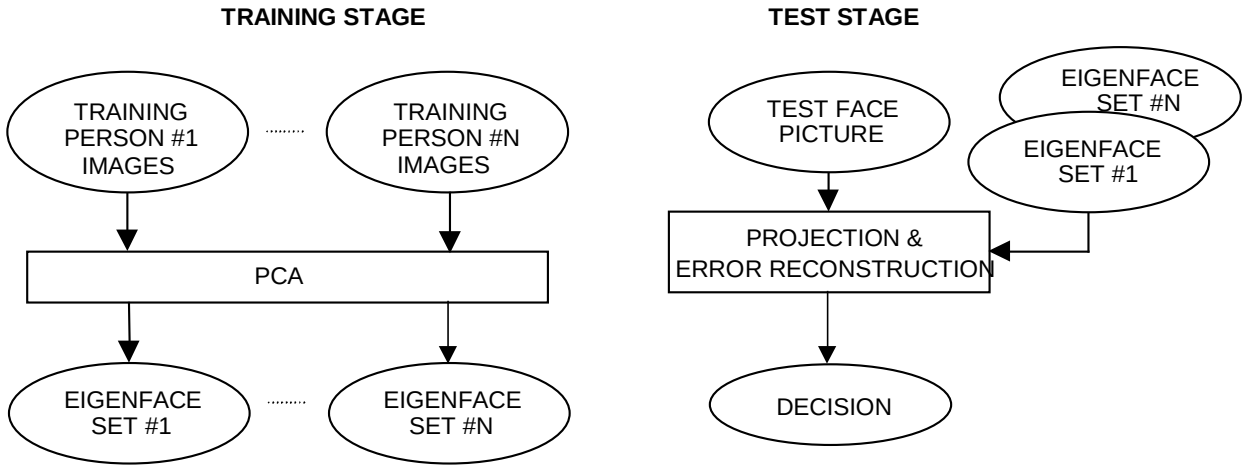

Figure 1. Block diagram of the self-eigen approach for face recognition

3. Results

We have performed different video sequence recognition tests using the CIF MPEG-7 test sequences. We have selected manually five faces of two different persons. These faces have been selected from the video sequence to be indexed. In some other cases these faces may be obtained from a training set which already contains the images to be indexed. Each set of five faces have been mirrored in order to obtain more views of the same person. This means that the training set for each person contains ten faces. Then, two different PCA have been calculated, one over each training set. As a result of the PCA, two different sets of eigenfaces have been obtained.

To check the recognition system, the test images have been extracted from the video sequence. Each face to be recognized has been projected and reconstructed using each set of the different eigenfaces. The minimum reconstruction error is selected and if this value is smaller than a given threshold, then the face with the minimum reconstruction error is said to match the training image which generated the corresponding set of eigenfaces. If the reconstruction error is bigger than the threshold then the image is said to be unknown which means that it does not match either of the faces of the training images.

Figure 2 shows two sets of training images selected from the video sequence to be indexed. Figure 3 shows some examples of recognized images in CIF format. Please notice that although the size, the orientation and the expresion of the test images are quite different from the training images, the system is still able to recognize them. In more than 90% of the cases, the system has been able to identify the correct person in the corresponding video frame, which is a high success rate for video indexing applications. When increasing the number of persons to be found, the results are very similar.

Figure 2. Original training images (real size)

Figure 3. Recognized test images (CIF)

Figure 3. Recognized test images (CIF)

References

[1] L. Lorente, L. Torres, "*Face recognition of video sequences in a MPEG-7 context using a globa eigen approach*", International Conference on Image Processing, Kobe, Japan, October 25-29, 1999.

[2] 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.

[3] 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.
[4] ISO/IEC JTC1/SC29/WG11. MPEG Requirements Group. "*MPEG-7: Context and Objectives*", Doc. ISO/MPEG N2460, October 1998 / Atlantic City, USA.

[5] 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.

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

[7] L.Torres, Ferran Marqués, Luis Lorente, Verónica Vilaplana, "*Face location and recognition for video indexing in the Hypermedia projec*t", European Conference on Multimedia Applications, Services and Techniques, Madrid, Spain, May 26-28, 1999.

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