PR 1463
pp: 1--11 (col.fig.: Nil)
PROD. TYPE: TYP
ED: Indira
PAGN: RDM – SCAN: Mangala

PERGAMON

Pattern Recognition 000 (2001) 000–000

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Automatic face recognition for video indexing applications☆

Luis Torres*, Josep Vilà

Department of Signal Theory and Communications, Polytechnic University of Catalonia, 08034 Barcelona, Spain

Received 16 November 2000; accepted 16 November 2000

Abstract

This paper presents an efficient automatic face recognition scheme useful for video indexing applications. In particular the following problem is addressed: given a set of known face images and given a complex video sequence to be indexed, find where the corresponding faces appear in the shots of the 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. *Conventional* face recognition schemes are not well suited for this application and alternative and more efficient schemes have to be developed. In this paper, in the context of Principal Component Analysis for face recognition, the concept of self-eigenfaces is introduced. In addition, the color information is also incorporated in the face recognition stage. The face recognition scheme is used in combination with an automatic face detection scheme which makes the overall approach highly useful. The resulting scheme is very efficient to find specific face images and to cope with the different face conditions present in a complex video sequence. Results are presented using the test sequences accepted in the MPEG-7 video content sequences set.

1. Introduction

Face recognition has been object of much interest in the last years [1,2]. 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, an increasing amount of audio-visual material is becoming available in digital form in more and more places around the world. With the increasing availability of potentially interesting material, the problem of identifying multimedia information is becoming more difficult. As a result, 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, among others, within the activities of the new standard MPEG-7 (multimedia content description interface) [3].

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 who searches a data base to answer the following type of query: Is there any face in this video sequence that matches that of Marlon Brando? More specifically, we are interested to know in our particular application, whether or not a specific face appears in a particular shot of the video sequence. The automatic answer to this problem is at this time very difficult, and it needs, at least, three stages: segmentation of the sequence in different shots, localization of objects that correspond to human faces within each shot and recognition of the faces. Fig. 1 shows the steps needed to recognize a face in a shot of a video sequence.

The first step relates to the segmentation of the video sequence in shots. The second step finds, within each shot, in which frames and in which positions the unknown faces are located. The output of this stage is a rectangular

☆This work has been partially supported by Grant TIC98-0422 of the Spanish Government and by the European ACTS AC-361 Hypermedia project.

*Corresponding author. Tel.: +34-93-401-6449; fax: +34-93-401-6447.

E-mail address: luis@gps.tsc.upc.es (L. Torres).

0031-3203/01/$20.00 © 2001 Published by Elsevier Science Ltd on behalf of Pattern Recognition Society.
PII: S0031-3203(01)00064-4

Fig. 1. Face recognition steps.

framed face, as shown in Fig. 1, which has to be recognized in the recognition stage. The third step performs the matching of the unknown faces against the faces contained in a training images data base. In this paper we are mainly concerned with the recognition part although it is clear that this stage has to rely on good video shot segmentation and face localization approaches.

Refs. [4,5] provide adequate bibliography in video sequence shot segmentation techniques while Refs. [6–8] present references for face detection approaches. The technique explained in Refs. [9,10] has been used in our system. More details will be provided in the sequel.

Almost all efforts in face recognition have been devoted to recognize still images. A very few works have presented results on video sequences [11,12]. A combined face detection and recognition generic scheme has already been presented in Ref. [13]. In addition very few works address the problem of face recognition in complex video sequences. 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 many applications the test sequences are in standard compressed formats, such as MPEG-1 or MPEG-2, which poses additional problems. It is in this context that there is a need to develop efficient face recognition schemes which may take into account the different face conditions present in video sequences and the lower quality present in compressed sequences.

Face recognition approaches can be roughly divided into: geometric, template matching and transform techniques [1,2]. In the first one, geometric characteristics of the faces to be matched are compared [14,15]. This technique provides limited results although it was used extensively some years ago. Template matching represents an improvement on the geometric approach at the expenses of being slow [16,17]. Finally, transform approaches offer good performance at a reasonable recognition time [18,19]. The well known principal component analysis (PCA) fall under this umbrella and the associated techniques have been widely used for face recognition [18,20].

In the following, a proposal to handle the problem of face recognition in video sequences based on the PCA approach, in a video indexing application is presented. The PCA has been modified to cope with the problems that arise in video indexing applications. The eigenfaces introduced in the PCA are extended and improved by using the concept of self-eigenfaces. Then, the face recognition approach is combined with a face detection approach [9,10] to have a completely automatic face detection and recognition system.

Section 2 presents the basics of the PCA for face recognition. Section 3 will present the proposed approach and Section 4 will introduce two modifications to the basic approach, which improve the performance of the overall system. The first one is the introduction of the color information in the face recognition stage and the second one is the use of intraframe images in the face detection stage. Section 5 will present the results of the proposed face recognition system using the MPEG-7 video sequences set and finally Section 6 will draw some conclusions.

2. Principal component analysis for face recognition

Among the best possible known approaches for face recognition, PCA has been the object of much interest [18] and is considered as one of the techniques that provides the best performance [2]. In PCA, the recognition system is based on the representation of the face images using the so called eigenfaces. 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 root mean squares (RMS) sense. 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 face images of different persons similar to the unknown images which are to be recognized.

In the eigenface representation, every training face image is considered as a vector $\vec{x}$ of gray pixel values (i.e. the training images are rearranged using row ordering).

Fig. 2. Eigenfaces of a set of images of the Stirling data base.

Using these vectors, a good representation of the faces may be obtained. If we suppose that we have N face training vectors, $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N$, which are the realization of a stochastic process $\vec{X}$, the average face can be approximated by

$$\vec{\mu} \approx \frac{1}{N}\sum_{i=1}^{N}\vec{x}_i \tag{2.1}$$

and it can be shown that the orthogonal vectors that optimally represent the distribution of the training face images in the RMS sense are given by the eigenvectors of the covariance matrix.

$$\underline{\underline{\Sigma}}_x \approx \frac{1}{N}\sum_{i=1}^{N}(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^{\mathrm{T}}. \tag{2.2}$$

The eigenvectors $\vec{e}_i$, are usually referred to as eigenfaces because they look like faces. A key feature of the eigenfaces is that they form an orthonormal basis, so it is very simple to compute the components of any face in the eigenface space.

Fig. 2 shows the first 32 eigenfaces of a set of images belonging to the Stirling data base [21], where the first image is the average image. The eigenfaces have been ordered according to the corresponding eigenvalue. To obtain these images, the normalization stage used in Ref. [11] has been used.

It can be noticed that any training image can be obtained, without error, by a linear combination of the eigenfaces:

$$\vec{x} = \vec{\mu} + \sum_{i=1}^{N}\hat{x}_i\vec{e}_i, \tag{2.3}$$

where $\hat{x}_i = \vec{x}\cdot\vec{e}_i$.

The eigenfaces can also be used to represent the test faces to be identified. This is done by projecting the test faces on the eigenfaces. The first eigenfaces contain the most information, in the RMS sense, so that we can express approximately a test vector image $\vec{y}$ in the eigenspace in terms of the principal components of the training vectors:

$$\vec{y} \cong \vec{\mu} + \sum_{i=1}^{N}\hat{y}_i\vec{e}_i, \tag{2.4}$$

where $\hat{y}_i = \vec{y}\cdot\vec{e}_i$.

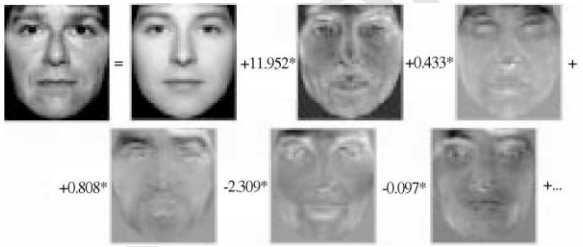

Fig. 3. Example of face image reconstructed using eigenfaces.

Any training face image which has been used to generate these eigenfaces can be perfectly reconstructed. A test face image, which is not in the data base, can be reconstructed up to a certain error using these eigenfaces. This concept will be used extensively in Section 3. Fig. 3 shows the representation of a face image using the eigenfaces, where the coefficients of the expansion have been obtained using Eq. (2.4).

The recognition is performed using the maximum likelihood principle by a distance computation. The selected training image is the one which has the minimal distance under the eigenbasis:

$$n_0 = \arg\min_{1\leqslant i\leqslant N} d(\vec{x}_i, \vec{y}), \tag{2.5}$$

where N is the number of face training images.

The test face is thus matched to the training face whose eigen representation is the most similar. The Euclidean distance or the Mahalanobis distance are commonly used [11]. The eigenface concept can be extended to any desired eigenfeature (mouth, nose, eyes, etc.) [20]. Half eigenfaces, called eigensides have also proved to be a useful tool for face recognition when part of the face is distorted or not available [11].

One of the main problems posed by the PCA when used for face recognition applications is the enormous size of the covariance matrix. As an example, for training images of 128×128 the covariance matrix reaches a size of $128^2 \times 128^2$ which becomes highly impractical. This problem can be solved using a singular value decomposition which allows to find the corresponding eigenvectors in a lower dimensional space [22].

3. Face recognition using self-eigenfaces

In personal identification applications, 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 on a set of different views of the same face 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 what is wanted is to find out whether or not three specific persons are in the sequence. The training images consist of different views of the same persons. A different PCA is performed on each set of views of each person giving three different sets of eigenfaces, one for each of the persons who must be recognized. It can be noticed that the main difference with the normal eigenface approach [18] is the number of different sets of eigenfaces and the type of training face images used.

The test and decision stages have to be modified accordingly. In the self-eigenface approach, each test image to be recognized is projected and reconstructed using each one of the sets of the different eigenfaces. In the example above, for each test face image to be recognized three different reconstructions are found, one for each set of eigenfaces. The unknown images will be said to match a particular face, when the corresponding reconstruction error using a set of eigenfaces be minimum. This decision stage relies on the fact that the set of self-eigenfaces obtained from the views of the same person are a good approximation of the base of the subspace of all the views of that face. Fig. 4 shows a simplified block diagram of the proposed approach. The use of an appropriate threshold may also tell if the test image does not match any of the training images.

An important step in the recognition process is the image normalization approach used to minimize the differences due to changes in size, expression and orientation of the training and the test image set. The normalization process can be manual or automatic. In a manual approach, representative points of the image to be normalized are manually extracted and the face is mapped against a predefined model. An example of a manual normalization was presented in Ref. [11], where the Candide image model [23] was used. However, although this technique provides very good results, it is very time consuming and highly impractical which is unacceptable in video indexing applications. Therefore, a very simple normalization technique has been used which improves the speed of the overall scheme at the expenses of recognition performance. The normalization stage defines one standard face with fixed height and width. Then the training and test images are resampled to fit these measures using a bilinear interpolation. A histogram normalization is also applied to the original images to enhance contrast. Notice that, before normalization, the test images correspond to the rectangular framed faces obtained in the face detection process (Fig. 1).

In order to show the quality of the reconstructed images using self-eigenfaces, Fig. 5 shows a view of the original test image Ana extracted from the *news 11* MPEG-7 video test sequence. All the MPEG-7 test sequences and much of the existing material to be indexed are available in compressed form which poses additional difficulties to

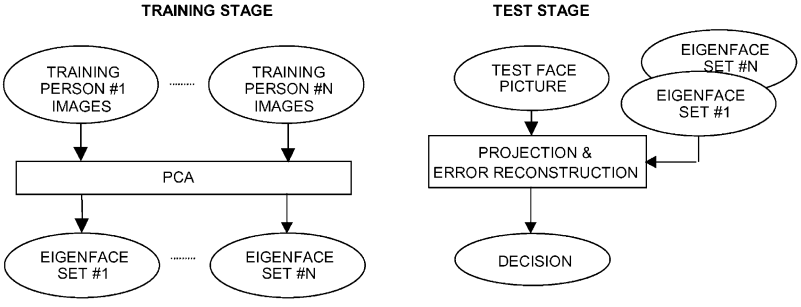

Fig. 4. Block diagram of the self-eigenface approach for face recognition.

Fig. 5. Original image Ana (CIF: 352 × 288).

the recognition problem. Fig. 6 shows five training images extracted of the same sequence and Fig. 7 shows the same images after normalization and mirroring. The mirroring effect, which relies on face symmetry, tries to improve the quality of the self-eigenfaces by providing additional variations to the training images. Fig. 8 presents the self-eigenfaces of Ana using the 10 training images of Fig. 7. The self-eigenfaces have been ordered according to the corresponding eigenvalues. It can be observed that the eigenfaces are of lower quality, when compared to the eigenfaces of Fig. 2, due mainly to the very simple image normalization process used.

Fig. 9. Left: face in Fig. 5 reconstructed with the self-eigenfaces of Ana and error image. Right: face in Fig. 5 reconstructed with the self-eigenfaces of María and error image.

It is clear that any of the images of Fig. 7 can be perfectly reconstructed using the self-eigenfaces of Fig. 8. However, any other view of Ana reconstructed with the same self-eigenfaces will give a reconstruction error. Fig. 9 (left) shows Fig. 5 reconstructed with its own self-eigenfaces and the error image. Fig. 9 (right) also shows Fig. 5 reconstructed with the self-eigenfaces of the image María, whose training normalized images are shown in Fig. 10, and the corresponding reconstruction error. It can be observed that the reconstructed image with its own self-eigenfaces keeps very well the features of the original image, while the image reconstructed with the self-eigenfaces of María keeps the features of that image. This fact is the key point in our face recognition approach.

Fig. 6. Training images of Ana.

Fig. 7. Normalized and mirrored images of Ana (image size: 50 × 70).

Fig. 8. Self-eigenfaces of Ana.

Fig. 10. Normalized and mirrored images of María.

4. Improvements of the self-eigenface approach

All the development and corresponding results of the previous section have been found using black and white images. In the same context, all the training and test images were MPEG-1 encoded. Improvements of the basic self-eigenface approach which incorporates the color information and some features of the bit-stream of standard compressed video sequences are presented now.

4.1. The use of color information

A common feature found in practically all technical approaches proposed for face recognition is the use of only the luminance information associated to the face image. One may wonder if this is due to the low importance of the color information in face recognition or due to other less technical reasons such as the no wide availability of color image data bases. However, some experiments performed in a eigen based approach to face recognition show that the use of the color information improves the overall performance of the scheme [24].

Based on these results, the self-eigenface approach explained above has been improved by using the color information. Several possible color spaces may be used. A variety of tests were done to select the best color space among the RGB, YUV and HSV ones [24]. Experimental results show that the YUV color space provides the best performance in terms of recognition ratio when compared against the RGB and HSV color spaces. An average increase of 4% in recognition ratio with respect to the use of only the luminance information can be experimentally established for different image data bases [24]. We note in passing that the YUV color space is widely used in coding applications which is very useful in the case that the video sequences are compressed in standard formats as this the case for MPEG-7 test video sequences.

The self-eigenface approach has been modified such that a different PCA is performed on each set of Y, U and V components of each person. To clarify the process, let us use the same example of Section 3. Assume that a video sequence is to be indexed and what is wanted is to find out whether or not three specific persons are in the sequence. The training images consist of different views of the same persons. Each view is first decomposed in its YUV components. Then a different PCA is performed on each set of components of each person giving nine different sets of eigenfaces, three for each of the persons who must be recognized (one for the Y component, one for the U component and one for the V component). The normalization process is applied to each component.

The test stage has to be modified to take into account the importance of each of the components. Each Y, U and V component of each test image is reconstructed using the corresponding Y, U and V self-eigenfaces found in the training stage. The composed reconstruction error (CRE) of an image is found through

$$CRE = \frac{1}{\omega_Y + \omega_U + \omega_V}(\omega_Y \delta_{eY} + \omega_U \delta_{eU} + \omega_V \delta_{eV}) \tag{4.1}$$

where ω_X is the weight of each component and δ_{eX} is the reconstruction error of each one of the components. The importance of each weight is related to the importance of each color component. In our case, the weights have been found empirically as no mathematical relationship seems to exist which may provide the right weight for each component. In the results section, specific values will be provided for these weights.

4.2. The use of intraframe images in MPEG-1 encoded video sequences

In order to recognize a face, our combined detection and recognition scheme should extract and recognize a face in each frame of the video sequence. In order to speed up the detection stage [9,10], only one out of M images are processed within a given shot. The test recognition stage is then only applied to these images. Much of the video material to be indexed, as is also the case for the MPEG-7 test sequences, is in standard compressed format which uses the concept of intraframe and interframes images. This implies that some images are of good quality (I frames) while others may be of lower quality (P or B frames). Fig. 11 shows frames 595–597 (B-I-B) of the *news 11* MPEG-7 test sequence. It can be observed the difference in quality between the different frames and that the image with the best visual quality corresponds to that of the I frame.

In all our experiments, the detection and test recognition stage has been performed on the I frames of the encoded sequences. It is clear that for quiet video images this improvement will not be noticeable while for

Fig. 11. Frames 595–597 (B-I-B) of the news 11 MPEG-7 test sequence.

Table 1
Video sequences used in the experiments

Video sequence	News 11	News 12	Contesting
Content	News of Spanish TV (TVE)	Weekly news (TVE)	Contest "Saber y ganar"
Duration	28′33″	18′16″	15′01″
Number of frames	42.802	27.383	22.518
Number of shots	268	99	135
Frame rate	25	25	25
Size	CIF (352 × 288)	CIF (352 × 288)	CIF (352 × 288)

images presenting a fast motion, the improvement may be considerable.

5. Results

In order to prove the validity of the proposed self-eigenface approach for automatic face detection and recognition, the scheme has been applied over a variety of MPEG-7 test video sequences. These video sequences are of moderate complexity and appropriate for video indexing applications. Table 1 provides information on the video sequences used in the experiments.

We recall that our application tries to detect whether or not a particular person appears in a given shot of a given video sequence. To evaluate the system, 12 persons appearing in the test video sequences have been looked for. The number of different persons appearing in the video sequences is much bigger than 12. Fig. 12 shows the persons used in the experiments.

Table 2 presents the number of show ups of each person in the shots of each video sequence. Persons have been numbered from left to right and top to bottom according to Fig. 12.

5.1. Training stage

To form the training set, five different views of the face of each person have been selected and extracted manually from the video sequence to be indexed. Thus, a set of rectangular framed faces of views of the training images have been obtained. A very complete graphic interface has been designed to that effect [25]. These training faces may be also obtained from another video sequence which contains the images to be indexed. In order to form a subspace that reconstructs the test image with the minimum reconstruction error, the views of each person should be as different as possible. Each training face is mirrored in order to obtain more views of the same person. After this mirroring process the training set of each person contains 10 faces.

Table 2
Number of show ups per person in each video sequence

Persons	Video sequence	Show ups/ number of shots
Person 1	News11.mpg	24/268
Person 2	News11.mpg	12/268
Person 3	News11.mpg	13/268
Person 4	News11.mpg	5/268
Person 5	News11.mpg	5/268
Person 6	News12.mpg	3/99
Person 7	News12.mpg	2/99
Person 8	News12.mpg	3/99
Person 9	Contesting.mpg	40/135
Person 10	Contesting.mpg	33/135
Person 11	Contesting.mpg	32/135
Person 12	Contesting.mpg	28/135
Total show ups		200/2177

The normalization stage changes automatically the size of all training images to a fix size of 50 × 70. The size of the normalized images is not a critical parameter. However, a very small size (25 × 35, for instance) would render the face recognition stage useless.

Once all the training images have been found and normalized, a PCA is performed on each set of Y, U and V

Fig. 12. Persons used in the experiments.

components of each person as explained in Section 4.1. Once the set of different eigenfaces is found for each training image the system is ready to enter into the face recognition stage. Notice that all the training stage can be done off-line.

5.2. Recognition stage

As we are interested to know if a particular person is in a given video sequence and in a given shot, in the recognition stage the video sequence is first divided into shots. To find the shots the technique presented in Ref. [26] has been used. This technique is based on changes detected in the image histogram and has proved to be very reliable. In order to have all the shots detected correctly, a manual validation is carried out after the automatic shot detection process. Then, the automatic face detection scheme presented in Ref. [10] is applied to I frames of the video sequences. In order to have a completely automatic face detection and recognition system, no manual validation has been used after the detection scheme. Once detected, the faces are normalized. Notice that an incorrect face detection will produce a failure in the recognition stage.

Each Y, U and V component of each test face obtained in the face detection scheme is projected and reconstructed using each set of the corresponding eigenfaces. The minimum composed reconstruction error is found according to Eq. (4.1) and if this value is smaller than a given threshold, then the face with the minimum composed 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.

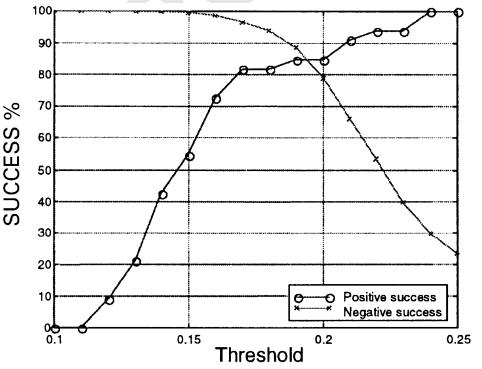

Fig. 13. Positive and negative successes of the system.

A critical point is the selection of the composed reconstruction error threshold. As we have not been able to find this value from any mathematical development, the threshold has been selected by empirical methods. In order to select it, the results of the complete face recognition system have been divided into positive success and negative success. A success is called positive when the system has decided correctly that the seeked person is in a specific shot. A success is called negative when the system has decided correctly that the seeked person is not in a specific shot. Fig. 13 shows a graph of the positive successes and negative successes obtained when the

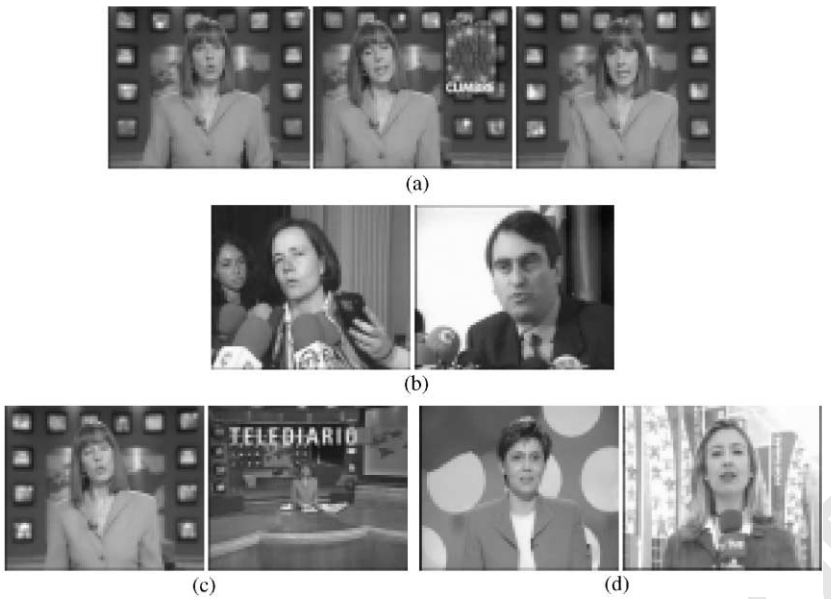

Fig. 14. Results corresponding to person 1: (a) positive success; (b) negative success; (c) recognition failure; (d) erroneous matching to person 1.

face recognition system has been applied to detect if all 12 persons were in all the video sequences.

The threshold has been selected at the intersection of the positive successes and negative successes which is 0.19. This selection implies that the same importance is given to each type of success which makes sense in the selected application. Other parameters which have to be selected are the weights given to each of the Y, U and V component. Our experiments show that variations in these weights do not influence very much the performance of the whole system. A good compromise has been found by selecting $\omega_Y = 1$, $\omega_U = 0.8$ and $\omega_V = 1.25$. Using these parameters, the percentage of success of the system is of 84.85%. Notice that this percentage corresponds to the same number of positive and negative successes. For positive successes, this percentage is quite high, given the complexity of the video sequences. In the event that the detection system fails to detect a face, the recognition system can still give an elevate number of negative successes. This might be the case, for instance, when the image contains some color which may be similar to that of the human skin (furniture, etc.). In this case the image which is presented to the recognition system is not a face and will not be matched to the face which is being seeked in the shot. This will cause a negative success. As examples, Fig. 14 shows some results corresponding to person 1. Fig. 14(a) shows good recognitions of person 1 (positive success) and Fig. 14(b) shows images which have not been correctly matched to person 1 (negative success). Fig. 14(c) presents failures of the system while Fig. 14(d) shows images which have been matched erroneously to person 1. Fig. 15 presents the same type of results referred to person 10. Some failures, are due to the very small size of the face image as in Fig. 14(c). Other failures are due to characteristics of the test images (glasses, occluded face), which converts the face recognition task in a very difficult problem, as in Fig. 15(d).

Notice that in spite of the size, orientation and expression of the test images, the system is still able to recognize most of them. This is mainly due to the correct performance of the face detection scheme and due to the different view conditions of the same image selected in the training process. In 85% of the cases, the system has been able to identify the correct person in the corresponding shot, which is a high success rate for face recognition in video sequences due the uncontrolled types of faces found in these sequences.

We recall that the results have been found using the face detection and face recognition schemes combined. Our experiments show that the detection system influences very much the overall scheme. As a very simple example, Fig. 16 shows an erroneous face detection result for person 9 which affects tremendously to the recognition part. For this particular case, it has been verified that if the face detection scheme had provided the adequate result (in the form of a rectangular framed face), the face recognition scheme would have recognized person 9. In some cases in spite of the failure of the detection stage, the recognition stage would also fail. Fig. 17 is an example of this situation. The detection system fails to provide a rectangular framed face (background is added to the face) but in the case of a good detection, the face

Fig. 15. Results corresponding to person 10: (a) positive success; (b) negative success; (c) recognition failure; (d) erroneous matching to person 10.

Fig. 16. Person 9 and a bad face detection (good recognition).

Fig. 17. Person 9 and a bad face detection and bad recognition.

recognition stage would not have been able to recognize the face either due to the very rotated test image. It has to be emphasized, that the detection face scheme is based on detecting frontal face views which is one of the main reasons of failure in the presence of very rotated images.

It can be generally stated that a failure in the detection stage causes an increase in negative successes and a decrease in positive successes. As the objective of this paper is to present the self-eigenface approach, a detailed study of the face detection system and its influence on the overall scheme is beyond the scope here. For more details see Refs. [9,10].

6. Conclusions

This paper has presented a self-eigenface approach to the problem of face recognition in a video indexing application. The scheme has been used in combination with an automatic face detection approach and has shown good results for moderate complex video sequences. The use of the color information and I frames for MPEG-1 compressed video sequences have shown an improvement on the performance of the system.

Acknowledgements

The authors would like to thank Profs. Driss Aboutajdine and Josep Vidal for the organization of the International Symposium on Image/video Communication over Fixed and Mobile Networks, held in Rabat, Morocco, April 17–20, 2000 in the context of the project "*Coding and transmission of digital video for MPEG-4 and for mobile communications of third generation UMTS*" of the Spanish and Moroccan cooperation program. The authors would like also to thank Ferran Marqués and Philippe Salembier for their useful comments which have improved the overall quality of this paper.

References

[1] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (5) (1995) 705–740.

[2] J. Zhang, Y. Yan, M. Lades, Face recognition: eigenface, elastic matching, and neural nets, Proc. IEEE 85 (9) (1997) 1423–1435.

[3] ISO/IEC JTC1/SC29/WG11, MPEG Requirements Group, MPEG-7: Overview, Doc. ISO/MPEG N3445, Geneva, May/June 2000.

[4] A. Hanjalic, H.J. Zhang, Optimal shot boundary detection based on robust statistical models, Proceedings of the IEEE Multimedia Systems'99, Vol. 2, Florence, June 1999, pp. 710–714.

[5] J.S. Boreczsky, L.A. Rowe, Comparison of video shot boundary detection techniques, Proceedings of the SPIE Conference on Storage and Retrieval for Still Image and Video Databases IV 2670 (1996) 170–179.

[6] H. Wang, S.-F. Chang, A highly efficient system for automatic face region detection in MPEG video, IEEE Transactions on Circuits and System for Video Technology 7 (4) (1997) 13.

[7] M.H. Yang, N. Ahuja, Detecting human faces in color images, IEEE International Conference on Image Processing, Chicago, IL, October 4–7, 1998.

[8] C. García, G. Tziritas, Face detection using quantized skin color regions, merging and wavelet packet analysis, IEEE Trans. Multimedia 1 (3) (1999) 264–277.

[9] V. Vilaplana, F. Marqués, P. Salembier, L. Garrido, Region-based segmentation and tracking of human faces, European Signal Processing Conference, EUSIPCO-98, Vol. I, Rhodes, 1998, pp. 311–315.

[10] F. Marquès, V. Vilaplana, Face segmentation and tracking based on connected operators and partition projection, Pattern Recognition, this issue.

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

[12] S. McKenna, S. Gong, Y. Raja, Face recognition in dynamics scenes, British Machine Vision Conference, 1997.

[13] L. Torres, F. Marqués, L. Lorente, V. Vilaplana, Face location and recognition for video indexing in the Hypermedia project, European Conference on Multimedia Applications, Services and Techniques, Madrid, Spain, May 26–28, 1999.

[14] W. Bledsoe, The model method in facial recognition, Panoramic Research Inc., Technical Report PRI:15, Palo Alto, CA, 1964.

[15] A.J. Goldstein, L.D. Harmon, A.B. Lesk, Identification of human faces, Proc. IEEE 59 (5) (1971) 748–760.

[16] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (10) (1993) 1042–1052.

[17] P. Kruizinga, N. Petkov, Optical flow applied to person identification, Department of Mathematics and Computer Science, Rijksuniversiteit Groningen, 1994.

[18] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, 1991, pp. 586–591.

[19] L. Wiskott, J.M. Fellous, N. Krüger, C.v.d. Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.

[20] A.P. Pentland, B. Moghaddam, T. Starner, M. Turk, View-based and modular eigenspaces for face recognition, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994, pp. 84–91.

[21] University of Stirling face database, http://pics.psych.stir.ac.uk.

[22] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Am. 4 (3) (1987) 519–524.

[23] Image Coding Group, The Candide software package, Linköping, University, 1994.

[24] L. Torres, J.Y. Reutter, L. Lorente, The importance of the color information in face recognition, International Conference on Image Processing, Kobe, Japan, October 25–29, 1999.

[25] Hypermedia Projects, ACTS AC-361, Continuous audiovisual market in Europe, June 1998–May 2000.

[26] J. Bescós, J. Menéndez, G. Cisneros, J. Cabrera, J. Martínez, A unified approach to gradual shot transition detection, International Conference on Image Processing. Vancouver, Canada, September 10–13, 2000.

About the Author—LUIS TORRES received the degree of Telecommunication Engineer from the Telecommunication School of the Polytechnic University of Catalonia, Barcelona, Spain, in 1977, and the Ph.D. in Electrical Engineering from the University of Wyoming, USA, in 1986. He is currently an Associate Professor at the Polytechnic University of Catalonia. His main research interests are image and video coding, image and video analsis and face detection and recognition. Luis Torres has published actively in different conference and journal research papers. He is also editor along with Murat Kunt of the book "Video Coding: The Second Generation Approach", Kluwer Academic Publishers, January 1996.

Luis Torres has worked in 40 national and international projects in very low bit-rate video coding applications, face detection and recognition, MPEG-4 and MPEG-7 standardization activities. Luis Torres is currently serving as an Associate Editor for the IEEE Transactions on Image Processing, and will be the General Chair of the International Conference on Image Processing, ICIP, to be held in Barcelona, September 2003.

About the Author—JOSEP VILÁ received the degree of Telecommunication Engineer from the Telecommunication School of the Polytechnic University of Catalonia, Barcelona, Spain in 2000. He is currently finishing his degree in Electronic Engineering and is working in Indra-Espacio developing communication systems. His interests include image and signal processing, and digital communications.