AN AUTOMATIC FACE DETECTION AND RECOGNITION SYSTEM FOR VIDEO INDEXING APPLICATIONS*

Emiliano Acosta [1] *Luis Torres*[1] *Alberto Albiol*[2] *Edward Delp*[3]

(1) Polytechnic University of Catalonia (2) Polytechnic University of Valencia (3) Purdue University
Barcelona, Spain Valencia, Spain West Lafayette, USA
emiliano.acosta@upcnet.es luis@gps.tsc.upc.es alalbiol@dcom.upv.es ace@ecn.purdue.edu

ABSTRACT

The objective of this work is the integration and optimization of an automatic face detection and recognition system for video indexing applications. The system is composed of a face detection stage presented previously which provides good results maintaining a low computational cost. The recognition stage is based on the Principal Components Analysis (PCA) approach which has been modified to cope with the video indexing application. After the integration of the two stages, several improvements are proposed which increase the face detection and recognition rate and the overall performance of the system. Good results have been obtained using the MPEG-7 video content set used in the MPEG-7 evaluation group.

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, 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) [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 that browse a data base to answer the following type of query: Is there any face in this video sequence that matches that of Sean Connery? 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.

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 [1], [2]. In addition, in many data base applications the accepted test sequences are in MPEG-1 format what poses additional problems. There is a need then, to develop efficient face detection and recognition schemes which may take into account the CIF format and the low quality present in MPEG-1 sequences.

It is in this context that this paper presents a combined automatic face detection and recognition scheme for video indexing applications. Our face segmentation and detection scheme is based on a skin detection approach followed by segmentation and region grouping [4] which provides good results at low computational cost. Our face recognition scheme is based on Principal Component Analysis [7] which has been modified to cope with the video indexing application [5]. The objective of this paper is to provide some improvements to both the face detection and the face recognition scheme which increase the performance of the overall approach.

Section 2 presents a summary of the face detection algorithm and the proposed improvements. Section 3 presents the basics of the PCA scheme and a new algorithm which modifies the face recognition decision stage and increase the recognition rate. Finally Section 4 shows some results which prove the effectiveness of the modifications here proposed on the overall scheme.

* This work has been supported by the US-Spain Joint Commission for Scientific and Technological Cooperation.

2. FACE DETECTION STAGE

The face detection stage used in our system [4] is presented in Figure 1.

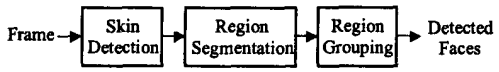


Figure 1. Detection algorithm scheme.

The skin pixel detection is performed with a simple colormap, using the YCbCr color space. The second block corresponds to the segmentation algorithm [6] which is performed in two stages, where the chrominance and luminance information is used consecutively. For each stage an algorithm which combines pixel and region based color segmentation techniques is used. After the segmentation, a set of connected homogenous skin-like regions is obtained. Then, potential face candidates (FC) are obtained by an iterative merging procedure using an adjacency criterion. Once the set of FC is built, it is necessary to remove the ones that do not match to any face. To that end some constrains regarding shape, size and overlapping are used.

Some results are shown in Figure 2. Notice that the algorithm is able to produce good face candidates but also provides some candidates which do not correspond to any face. Most of these erroneous face candidates will be not recognized in the face recognition stage and will be discarded, but at the expenses of increasing unnecessarily the computational cost. Thus, if there are many erroneous face candidates, the overall system becomes inefficient.

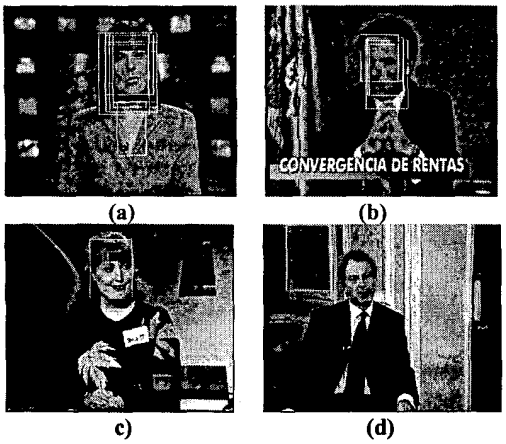


Figure 2. (a)(b) Images with many similar face candidates; (c) Face candidate that includes hair and (d) non-face candidate.

In order to eliminate erroneous face candidates, two new selection criteria of face candidates have been designed and are presented below.

2.1. Selection criterion based on color

Face areas are characterized to have an homogeneous chrominance component. Taking into account this fact, a new selection criterion han been designed in order to remove all the FC composed of regions whose average color differs substantially, as is the case when hair and face are included in the same candidate.

The new criterion is based on the chrominance component of the regions that form the candidate face regions. Given a FC composed of regions R_1, R_2, R_3 and whose average values of chrominance are $C_1=(cb_1,cr_1)$, $C_2=(cb_2,cr_2)$ and $C_3=(cb_3,cr_3)$ respectively, the distance between all the possible pairs of colors is obtained as follows:

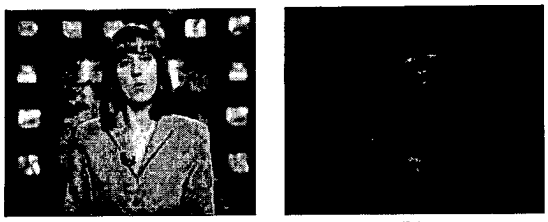
$$d(C_i,C_j) = \sum_i \sum_j (cb_i-cb_j)^2 - (cr_i-cr_j)^2$$

If the distance d(Ci, Cj) is greater than a certain threshold, the corresponding FC will be discarded. Using this very simple procedure, the results improve significantly and many erroneous FC are discarded.

2.2. Selection criterion based on texture

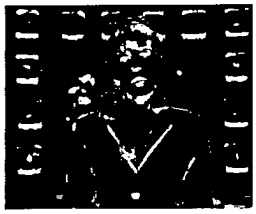
In many ocasions the background has a color very similar to that of a human face. In this case, the face detection algorithm may fail. In order to better discriminate between color like face and not face regions, a criterion based on texture has been introduced. The selection criterion is based on the detection of dark horizontal regions in the image (eyes, mouth, etc.). The objective is to remove all face candidates containing totally flat texture areas.

The first step is to apply a morphological horizontal erosion of size 3 in order to remove the brightness associated to white regions (eyes, teeth, etc.), (Figure 3 a). Then, a vertical close of size 15 is applied to eliminate the horizontal regions (Figure 3 b). The objective of this step is to obtain an image with horizontal zones. After applying a texture threshold, a binary open with a square structuring element of size 3 is performed to remove the spurious zones that still remain (Figure 3 c). Once the spurious zones are eliminated, the area of white pixels is calculated. If the area of each white spot is less than 3.5% of the total area of the FC or greater than 15%, the corresponding FC will be discarded.



(a) (b)

Figure 3. (a) Original grey scale image after erosion;(b) Vertical close of **(a).**



(c)

Figure 3. (c) Final image after binary open.

3. FACE RECOGNITION STAGE

The recognition stage integrated in our system is based on Principal Component Analysis (PCA) [7] which has been modified to cope with the video indexing application [5]. In particular, what is wanted here is to know whether a specific person is present in a particular shot of 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. 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 three specific persons appear 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 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 this 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 image will be said to match a particular face, when the corresponding reconstruction error using a set of eigenfaces be minimum. The use of an appropriate decision threshold may also tell if the test image does not match any of the training images.

This approach has been proved to be very efficient for the recognition of individuals providing high success rates [5]. However in some ocasions, this procedure is not robust and presents some deficiencies. In particular the use of a fixed decision threshold and the possibility that the reconstruction values of the many possible face candidates of the same frame lie close to the threshold, dificults the decision stage of the face recognition scheme. In order to cope with this situation and to improve the efficiency of the face recognition system, a new algorithm for the decision stage will be proposed in the next section.

3.1. Median Post-Processing

Let $\boldsymbol{Shot_x}$ be a vector whose components are the minimum reconstruction error of all face candidates of each one of the frames that composes the analyzed shot. That is:

$$Shot_x = [\, e_0, e_1, e_2, \ldots, e_{N-1}, e_N \,]$$

where x represents the number of the shot that is analyzed, N the number of frames that are processed within that shot and e the value of the minimum reconstruction error of each frame. For each shot, the median value of $Shot_x$ will be considered as the reconstruction error which will be compared against the decision threshold.

This approach allows that the decision criterion becomes independent from the number of processed frames of each shot, since very likely, positive shots (shots where the wanted person appears) will be represented by vectors with smaller component values than the negative ones. Therefore, the error that is in the central position of the vector will be smaller.

4. RESULTS

The face detection and recognition integrated system has been tested using the CIF MPEG-7 content set video sequences. Let us call *original system* the system that combines the face detection sheme presented in [4] and the face recognition scheme presented in [5]. Let us call *modified system* the one with the modifications presented in this paper.

Table 1 presents the number of face candidates found in two video sequences of the MPEG-7 video content set. The important reduction of face candidates can be noticed. This allows to eliminate non-significant face candidates which will not reach the recognition scheme.

Face candidates	Original System	Modified System
Contesting.mpg # frames: 22518	10.193	4.525
News11.mpg # frames: 42802	28.735	12.654

Table 1. Face candidates

The face recognition rate of the scheme with the proposed modifications presents an improvement of 5% against the *original system*. The recognition success rate when trying to recognize a person in a particular shot is now of 90% (positive and negative successes). That is, the number of times the system has been able to locate and recognize a person in the shot or to decide correctly that the person was not in the shot. The average running time (defined as the ratio between the executing time of the system and the analyzed sequence duration) has improved 30%. The total number of different persons analyzed is 12.

Figure 4. Faces automatically detected and recognized by the proposed system

5. CONCLUSIONS

An automatic face detection and recognition system has been presented. The system is based on a face detection and a face recogniton schemes proposed previously independently. The system is intended for video indexing applications and is able to tell whether or not a specific person is present in a video sequence shot. Recognition rates of 90% are achieved for a variety of MPEG-1 encoded sequences

6. REFERENCES

[1] 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.

[2] 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.

[3] ISO/IEC JTC1/SC29/WG11. "*Overview of the MPEG-7 Standard*", Doc. ISO/MPEG N4031, March 2001, Singapore.

[4] A. Albiol, L. Torres, C.A. Bouman and E. J. Delp, "*A simple and efficient face detection algorithm for video database applications*", Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, vol. 2, pp. 239-242, September 2000.

[5] L. Torres, L. Lorente and J. Vilà, "*Face recognition using self-eigenfaces*," Proceedings of the International Syposium on Image/Video Communications Over Fixed and Mobile Networks, Rabat, Morocco, pp. 44-47, April 2000.

[6] A. Albiol, L. Torres, E. Delp, "An unsupervised color image segmentation algorithm for face detection applications", IEEE International Conference on Image Processing, Thessaloniki, Greece, October 7-10, 2001.

[7] 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.