A Framework for Face Recognition from Video Sequences Using GWN and Eigenfeature Selection

Teófilo Emídio de Campos
Rogério Schmidt Feris
Roberto Marcondes Cesar Junior

{teo, rferis, cesar}@ime.usp.br
http://www.vision.ime.usp.br/~creativision
Departamento de Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo (USP),
Rua do Matão, 1010, PO-Box 66.281,
05315-970, São Paulo, SP, Brasil

Abstract. This work describes an ongoing project related to a new method to perform face recognition from video sequences. Faces are detected and tracked in a video sequence using Gabor wavelet networks. This process also allows locating and extracting facial feature regions around the eyes, nose and mouth. A modified Karhunen-Loève transform, defined with the aid of an automatic feature selection procedure, is used for feature extraction and face recognition. Several preliminary results are discussed.

Keywords: face detection, tracking of facial features, face recognition, Gabor Wavelet Networks, Principal Components Analysis and Feature Selection.

1 Introduction

Face recognition has emerged as an instigating research field both under the machine and the biological vision systems point of views. In fact, this research field has attracted intense and growing attention by the vision research community, finding many important practical applications, such as in security and human-machine interaction. This paper describes an ongoing project, including some preliminary results, about a face recognition system based on video sequences.

The methods for face recognition can be broadly grouped in two classes, i.e. static and dynamic approaches. The former concerns face recognition in single (static) images, while the latter concentrates on people possibly moving in video sequences. The static approach has developed many interesting and powerful techniques, succeeding in different applications. Nevertheless, as far as recognition in video sequences is concerned, much work still remains to be done [13].

Face recognition in video sequences often involves four important steps: (1) face detection; (2) face tracking along the video sequence; (3) feature extraction; and (4) recognition. It should be clear that the large amount of data involved in video sequences represent a challenge for real-time implementation of these four steps. In order to circumvent this problem, not all face regions will be taken into account by the foreseen system. Instead, the system is based on extracting information from important facial feature regions (FFRs) defined around specific landmarks, i.e. the eyes, the nose and the mouth.

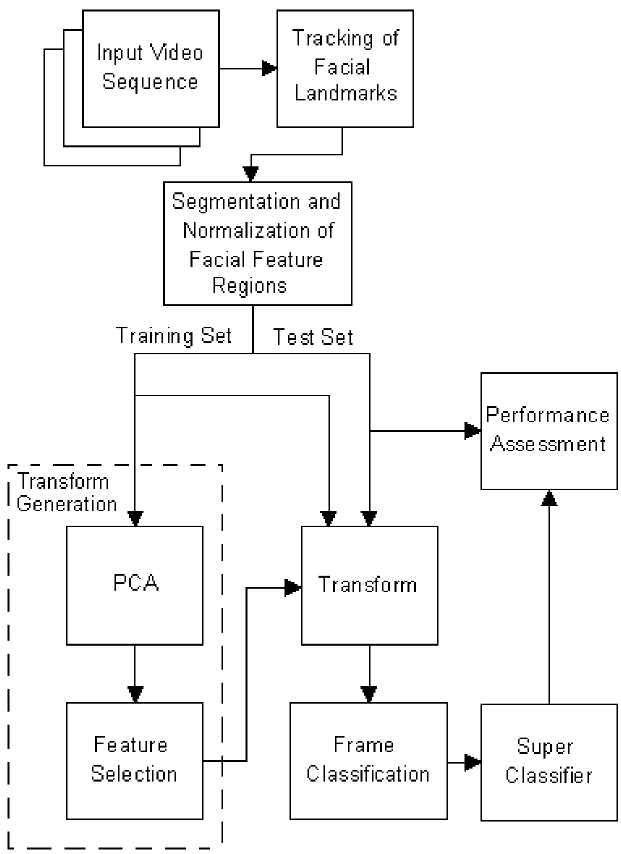

Figure 1 Overview of the project.

Figure 1 illustrates an overview of the system project. Using Gabor wavelet networks (GWN), the important facial landmarks (FLs) are detected and tracked, allowing the FFRs to be segmented and extracted. The features used for classification are extracted from the FFRs using the Karhunen-Loève transform. Nevertheless, instead of taking the eigenvectors (or eigenfeatures) associated to larger eigenvalues, an automatic feature selection algorithm will be applied in order to define a better transform from the training set, by choosing the most discriminative eigenfeatures. Once that the feature extraction transform has been defined from this procedure, it can be applied to the FFRs in order to generate the feature vectors that allow classifying the faces in each frame. Finally, a super classifier is supposed to collect the output classification results from each frame in order to recognize the face based on several classification results (one from each frame, in a voting scheme). All these steps are represented in Figure 1, which also includes a systematic performance assessment module that will be carried out in order to validate the system.

This work is organized as follows. Section 2 describes the GWN approach for face detection, tracking and location of the facial landmarks. Section 3 concentrates on feature extraction and selection. Finally, the face classification in each frame, the super classifier and performance assessment are topics presented in Section 4.

2 Detection and Tracking of Facial Features

The detection and tracking of facial features in video sequences is an important step of our approach, since it provides the normalized facial feature images, which are required by both the training and classification procedures. The topics concerned to this section can be divided in three parts: detection of faces, facial features and normalization of facial features. It is important to say that the user is supposed to move his head naturally while the camera acquires the image sequence, with non-controlled illumination conditions.

Face detection is performed by using a statistical skin-color model to segment face-candidates as well as a simple correlation procedure to verify the presence or absence of a face in each detected skin-color blob. Once a face is detected, its approximated scale and position are computed, and the skin-color region is converted into a gray-level image. These procedures are described in [1].

Facial features, i.e. eyes, nose and mouth, are then located and tracked along the video sequence. The method that we have used to accomplish this task is based on Gabor wavelet networks (GWNs) [2], an effective technique for object representation. Basically, the idea of GWN is to represent a face image as a linear combination of 2D Gabor wavelets, whose parameters (position, scale and orientation) are stored in the network nodes, while the linear coefficients are represented as the synaptical weights. The weights and wavelet parameters are determined optimally so that the maximum of image information is preserved for a given number of wavelets. This wavelet representation allows detection of facial features even in the presence of glasses, beard and different facial expressions. Feature tracking is robust to homogeneous illumination changes and affine deformations of the face image. Moreover, the approach considers the overall geometry of the face, thus being robust to deformations such as eye blinking and smile, which is usually a critical situation to most local-based traditional methods.

Figure 2 illustrates the face and facial feature detection, whereas figure 3 shows the tracking process in a specific video sequence[1]. The technique based on Gabor wavelet networks provides, in each frame, the orientation, scale, translation and shearing of facial features. Thus, using these parameters, we can segment and normalize each facial feature by applying a correct affine transformation in each frame. The resulting data are used by both the training procedure and the classification, as we will describe in the next sections.

Figure 2 Detection of face and facial features.

Figure 3 Tracking of facial features.

[1] http://www.ime.usp.br/~cesar/creativision/demo/

3 Feature Space Generation

In order to perform recognition, our system generates a feature space according to the method schematized in figure 4. Next two subsections contain further details about our methodology.

3.1 Eigenfeatures

After estimating the facial features locations, followed by segmentation and normalization of facial features regions from all the frames of the training set, the resultant images are used to train four principal components analysis (PCA) feature extractors (also called Karhunen-Loève extension), one for each facial feature, because PCA generates a representation better suited to classification procedures.

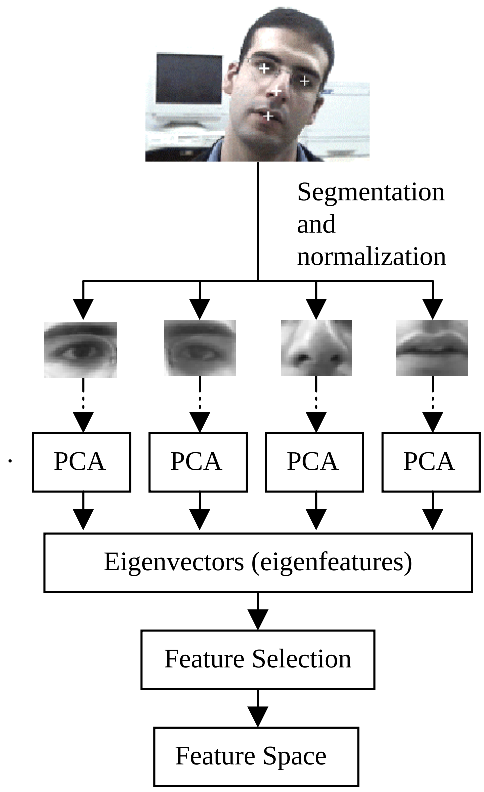

Figure 4 Feature space generation.

These PCA transforms are obtained through the method described by Turk and Pentland [3], but using only images of facial features instead of whole face images thus obtaining eigenfeatures. The terms eigenfeatures, eigeneyes, eigennoses and eigenmouths were created by Moghaddam and Pentland in [11]. The advantages of using only facial features includes computational cost, because smaller images require less processing to train the PCA; and recognition accuracy, due to the curse of dimensionality [9]. These advantages were reported in [4], and Brunelli and Poggio [7] also confirms these results for the template matching approach.

In order to perform dimensionality reduction, Turk and Pentland [3], simply select the first d eigenvectors of the covariance matrix of the training faces set ($d<m$, where m is the total number of eigenvectors). In this work, we apply a feature selection algorithm over a combination of all the eigenfeatures (eigenrighteyes, eigenlefteyes, eigennoses and eigenmouths).

Besides the well-know advantages of feature selection, like those discussed in [9], one of the motivations for performing automatic eigenfeature selection are the results obtained in [8], in which the performance of a PCA for face recognition system was improved when the first three eigenfaces were not used. The authors in [8] claim that there are some evidences that these eigenfaces are influenced by illumination changes, and not to inter-class variations. This fact provides evidence that it is possible to obtain better results by applying an automatic feature selection method to the eigenfeatures, instead of using the first d eigenvectors of each facial feature.

3.2 Feature Selection

Briefly, automatic feature selection is an optimization technique that, given a set of m features, attempts to select a subset of size d ($d<m$) that lads to the maximization of some criterion function, e. g. the distance among the training classes.

In [12], we described a feature selection strategy that produces good classification results without dependence of a specific classifier. These results are possible because that algorithm selects a feature set that maximizes distances among elements belonging to different classes and minimize distances among elements that belong to the same class. An advantage of that technique is that it is possible to obtain these results independently of the shape of distribution of the patterns in the feature space.

Briefly, the technique proposed in [12] is composed by the sequential floating search methods [5], using, as criterion function, the tolerance-based fuzzy distance [6]. The membership value of each pattern was defined as being inversely proportional to the distance between it and the prototype of its class. We defined the prototype of a class as the mean vector of the training samples of that class.

4 Classification

In order to perform classification of video sequences, the feature space is obtained by transforming the input video sequence using the PCA transform of section 3. Each individual frame is classified and a super-classifier is used to give the final classification result of the video sequence.

An evaluation procedure will be used to assess and improve the performance by determining which classifier is the best between the *K*-nearest-neighbor (*Knn*) and the minimum distance to the prototype. The evaluation system will also be used to determine the value of the parameter *K* for the *Knn* classifier. These classifiers are described in [10].

We use the voting scheme [9] as a super-classifier (combination method), but we plan to assess the performance of other super-classifiers, like those described by Jain et al. in [9].

Our ongoing work also includes performance assessment using the "leaving-one-out" strategy [10] and the mean of the correct classification rates.

We have obtained promising results applying the feature selection technique described in previous section over eye images and using a *Knn* classifier. Our preliminary tests were carried out using 174 images of eyes from 29 people, 6 images per person. The resolution of the images is 64 x 64 pixels. All images were used to train a principal components analysis system (PCA), obtaining 468 eigenvectors.

The ASF feature selection was done to select the best (according to the tolerance-based distance [6]) $d=30$ eigenfeatures. The obtained feature set was evaluated using the *Knn* classifier, for $K=3$, trained using all the images and testing using one image per person. The recognition accuracy was 97.7%, while, by using the first 30 eigenvectors (the traditional PCA classification method), the obtained recognition accuracy was 94.8%. This first result shows that our feature selection strategy is promising.

5 Concluding Remarks

We presented a new method to perform face recognition from video sequences. Our approach is based on: (1) detection using a statistical skin-color model to segment faces and GWN to detect facial landmarks; (2) tracking of facial landmarks using GWN and (3) recognition using eigenfeatures selection.

The main contributions of this work are the eigenfeatures selection for people recognition, and the association among GWN, PCA and feature selection using fuzzy distance.

As described before, the preliminary results of the detection and tracking systems as well as the results of the eigenfeature selection system are very promising. As future work, we plan to perform tests comparing the performance of eigenfeatures selection versus eigenfeatures without automatic feature selection. We also plan to compare it with other systems concerning face recognition from video sequences.

Acknowledgements

T. E. Campos is grateful to FAPESP for the financial support (99/01488-8), R. S. Feris is grateful to FAPESP (99/01487-1) and R. M. Cesar J. is grateful to FAPESP (98/07722-0) as well as to CNPq (300722/98-2).

References

[1] R. S. Feris, T. E. Campos and R. M. Cesar Jr.; "Detection and tracking of facial features in video sequences"; *Lecture Notes in Artificial Intelligence*, vol. 1793, pp. 127-135, Springer-Verlag, April 2000.

[2] V. Kruger and G. Sommer; "Affine real-time face tracking using a wavelet network"; Presented at the *ICCV'99 Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece, September 1999.

[3] M. A. Turk and A. P. Pentland; "Face recognition using eigenfaces"; *Proc. of the IEEE on Computer Soc. Conf.*, 1991.

[4] T. E. Campos, R. S. Feris, R. M. Cesar Jr.; "Eigenfaces versus Eigeneyes: First Steps Toward Performance Assessment of Representarions for Face Recognition"; *Lecture Notes in Artificial Intelligence*, vol. 1793, pp 197-206, Springer-Verlag press, April 2000.

[5] P. Pudil, J Novovicová and J. Kittler, “Floating search methods in feature selection”, *Pattern Recognition Letters* 15 (1994), pp. 1119-1125.

[6] R. Lowen and W. Peeters; “Distances between fuzzy sets representing gray level images”; *Fuzzy sets and Systems*, 99(2), pp. 135-150, 1998.

[7] R. Brunelli, T. Poggio; “Face recognition: features versus templates”; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10), pp. 1042-1052, October 1993.

[8] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman; “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 711-720, July 1997.

[9] A. Jain, R. P. W. Duin, J. Mao; “Statistical pattern recognition: a review”; *IEEE Transactions on Pattern Analysis and Machine Ingelligence*, 22(1), pp. 4-37, January, 2000.

[10] R. Duda, P. Hart, “Pattern Classification and Scene Analysis”, *Wiley,* New York, 1973.

[11] Baback Moghaddam, Alex Pentland; “Face Recognition using View-Based and Modular Eigenspaces”; Automatic Systems for the Identification and Inspection of Humans, *SPIE* Vol. 2277, July 1994.

[12] T. E. Campos, Isabelle Bloch, R. M. Cesar Jr.; “Feature Selection Based on Fuzzy Distances Between Clusters: First Results on Simulated Data”, *submitted to* Second International Conference on Advances in Pattern Recognition, Rio de Janeiro, Brazil, March, 2001. Preprint available at: *http://www.ime.usp.br/~cesar/creativision/.*

[13] R. Chellappa, C. L. Wilson and S. Sirohey; “Human and Machine Recognition of Faces: A Survey”; *Proceedings of the IEEE*, 83(5), pp. 703-740, May 1995.