Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

SaBP2.25

An Automated Tool for Face Recognition using Visual Attention and Active Shape Models Analysis

A. Faro, D. Giordano, C. Spampinato
Dipartimento di Ingegneria Informatica e Telecomunicazioni
University of Catania, Viale Andrea Doria 6
Catania 95125
ITALY
{afaro, dgiordan, cspampin}@diit.unict.it

Abstract – **An entirely automated approach for the recognition of the face of a people starting from her/his images is presented. The approach uses a computational attention module to find automatically the most relevant facial features using the Focus Of Attentions (FOAs). These features are used to build the model of a face during the learning phase and for recognition during the testing phase. The landmarking of the features is performed by applying the active contour model (ACM) technique, whereas the active shape model (ASM) is adopted for constructing a flexible model of the selected facial features. The advantages of this approach and opportunities for further improvements are discussed.**

I . INTRODUCTION

Over the last ten years or so, face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. Due to the nature of the problem, not only computer science researchers are interested in it, but neuroscientists and psychologists also. It is general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and viceversa. In fact, different studies in psychology and neuroscience are highly relevant for the design of these recognition systems. For example, discoveries in psychology of the importance of some features for recognition [8] have influenced the engineering literature [5]; on the other hand, automated systems provide a significant support for studies in psychology. For example, a study on the lighting from bottom of the face [11] has shown that if we use a lighting direction different by the one used normally, the recognition algorithm doesn't work.

A general statement of the face recognition problem (in computer vision) can be formulated as follows: given a video image of a scene, identify or verify one or more persons in the scene using a stored database of faces. The problem of face recognition is still open, since the proposed systems have evident shortcomings when compared to the human capability to recognize the faces. Thus, in the paper, we propose a biologically inspired approach for face recognition based on the simulation of visual attention, according to the model of visual attention proposed in [9]. This model allows us to detect the relevant facial features from an image (i.e. eyes, nose, mouth, etc.). For the recognition of such features many models have been proposed using specialized techniques designed to detect single facial features. Filtering techniques are successful [1], [15], [14] but they may not work well for some variation of the feature shapes. Methods based on deformable templates seems be more effective. Kass et al. [12] describe the use of Active Contour Models (ACM or "snakes") to contour an object in an image. Because snakes do not contain knowledge about expected shapes, this approach is not satisfactory because it's possible to detect other irrelevant structures present in the image. Yuille et al. [17] describe the use of deformable templates, based on simple geometrical shapes to locate eyes and mouths. The main problem of the Yuilles's models is that, for its applicability, the form of a given model should be sufficiently general and this is difficult to obtain in practice. There are many methods that use the active shape model (ASMs), [10] and [16], but they use a different approach for the detection of the features based on wavelet, Gabor filter, etc. Cootes et al. [13] have used an effective model for the interpretation and coding of face images with results in the range [70%-97%], but their approach uses a method where the landmarks are manually located to detect the main facial features.

To execute the face recognition process in real time or in a reasonable time interval, in this paper an entirely automated tool is proposed that suitably integrates some of the mentioned techniques. In particular, Sect.2 illustrates how the tool uses a visual attention module to identify, by a rough landmarking, the most significant facial zones, called features. Sect.3 shows how the optimal landmarking of such features is automatically performed by applying the ACM approach. In sect.4 the ASM technique is recalled to illustrate how the tool derives, during a learning phase, the models of the relevant facial features for each person to be recorded in the database. In the learning phase, the tool takes into account many images of the same person since the higher the number of images of the person, the higher the precision of the ASM technique. In the recognition phase, discussed in sect.5, the tool will compute how many features of the person under test match the features of the faces of the persons stored in our database. If such percentage, for some record of the database, is higher than a certain threshold, then the tool will signal the person has been recognized. The advantages and further improvements of this approach are discussed in the concluding remarks.

II. VISUAL ATTENTION AND FACIAL FEATURES

Many biological vision systems seem to apply serial computational strategy when inspecting complex visual scenes. Particular locations in scenes are selected based on their relevance from both the objective and subjective point of view, with reference to the observer. In a notable study of 1967, Yarbus demonstrated that perception of complex scenes involves complicated pattern of *fixations*, where the eye stands still, and *saccades*, where the eye moves to include in the fovea a part of the scene. Basically, fixations occur for zones that are salient to determine specific features of the scene under consideration.

The visual attention module proposed in this paper detects salient regions from a color image simulating saccades of human vision using the model proposed by Itti & Koch [9]. Inspired by psychological work, this module determines in a bottom-up fashion the conspicuities of different image properties, i.e., intensity, color and orientation. The intensity and the color maps are determined by mechanisms simulating the on-center-off-surround cells of the human visual system. Such mechanisms compute the intensity and color difference between image regions and their surroundings by using a linear filtering at 8 spatial scale. The orientation property is obtained by Gabor's filters.

An iterative lateral inhibition scheme instantiates for each feature map a competition for saliency. After competition is terminated, the property maps are combined into a single conspicuities map, one for each property. Then the conspicuities maps of the three mentioned properties are fused into a single saliency map by strengthening maps with few peaks before summing them up. Finally, the Focus of Attention (FOA) is directed to the most salient region.

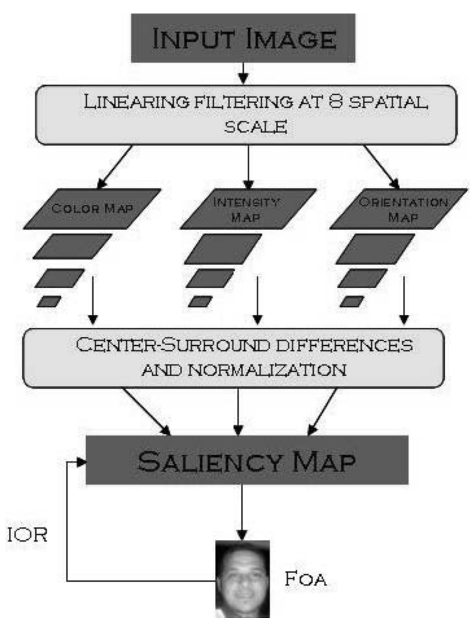

Figure 1: Bottom-up approach for modeling the visual attention process

Afterwards, this region is inhibited according to a mechanism called *inhibition of return (IOR)*, allowing the computation of the next FOA. The computed FOA is the starting point to identify the regions where the face features are located. In detail for this identification we use an approach based on a spatial analysis and filtering [6] to understand what region we are considering, e.g. if the point is located in the mouth, we use an analysis on the shape to select the minimum rectangle that contains this feature. The architecture of the visual attention module is shown in fig.1.

Examples of the calculus of saliency map and FOA for a face image are shown respectively in fig.2a and fig.2b. From this image we extract the zones corresponding to the FOA, thus obtaining the main features of the face, as shown in fig.2c. For each of the extracted features we compute, during a learning phase, a dynamic model that will be used for recognition, as described in the following section.

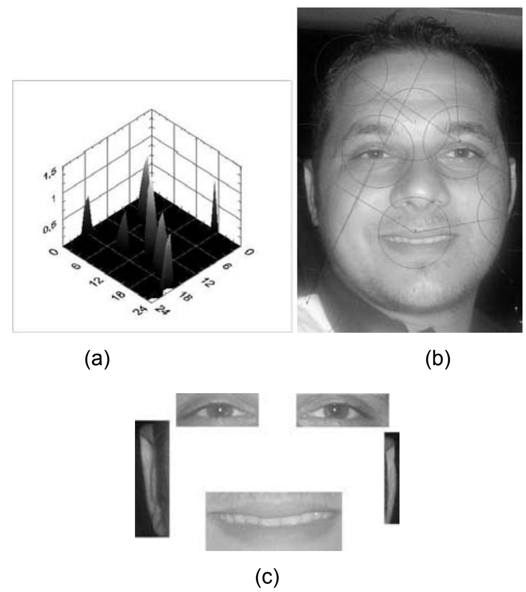

Figure 2: a) Saliency map, b) resulting FOAs after IOR and c) main facial features computed by the visual attention module

II. LANDMARKING THE FACIAL FEATURES BY ACTIVE CONTOUR MODEL

In the introduction we have pointed out that one of the main problems for determining a shape model is represented by the landmarking of the shape. For an automatic landmarking the proposed tool uses the ACM (Active Contour Models), often called "Snakes" [12] that are deformable models that "move" by following a criterium of minimum energy towards the salient features on an image (typically, the image contours). These features correspond to local minima of the ACM energy function to be used in the image processing.

Let us note that the snake is a low-level mechanism that seeks local minima rather than a global solution; it is therefore necessary to position the initial snake in proximity of the appropriate zone. In the tool the initial snake is the rectangle obtained by the FOA analysis illustrated in the previous section thus avoiding that it is attracted toward non desired zones of the image. In analytical terms, a snake is a parametric curve represented by the following parametric function:

$$v(s) = (x(s), y(s)) \text{ where } s \in [0, 1]$$

This curve is influenced by a set of internal and external forces defined as follows:

- Internal forces are those forces which characterize the snake, by giving it a greater or lesser ability to stretch and bend;
- External forces are determined by the image and are used to steer the model towards salient features of the image, such as light and dark regions and contours.

The total energy of the snake, i.e., E_{snake}, is defined in terms of two energy functionals as follows:

$$E_{snake} = \int_0^1 E_{\text{intern}}(v)\, ds + \int_0^1 E_{image}(v)\, ds$$

The internal energy E_{intern} contains two terms: one is controlled by the coefficient $\alpha(s)$, which represents elasticity, the other one is controlled by the coefficient $\beta(s)$, which represents rigidity. Concerning the external energy E_{image}, it is given by a continuous function of position designed to steer the snake towards contours. The final goal of the ACM technique is to find the parametric curve in the considered zone whose energy E_{snake} has the minimum value. The so obtained snakes are stable curves contouring the sought features. Figure 3 shows some steps of the automated landmarking process of the contour of a left eye by means of the mentioned ACM technique. The snake after 200 iterations (fig.3a) and especially the one after 1000 iterations (fig.3a) constitutes an optimal eye contour starting from which the ASM technique, recalled in the next section, may found the eye shape model.

(a)

(b)

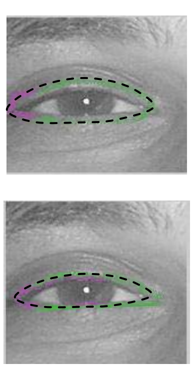

Figure 3: Snake of an eye image after 200 iterations (a) and after 1000 iterations (b)

III. MODELLING THE FACIAL FEATURES BY ACTIVE SHAPE MODEL

Several methods could be used to recognize the specific patterns represented by the high-level features extracted by the visual attention module. ASM (Active Shape Models) is a relatively recent technique that allows the construction of flexible models of shapes, able to capture the natural variability of a class of shapes. Generally, ASM is used for shape retrieval, i.e., to find in a given scene instances of the shape it represents. These shape models are created starting from a set of training shapes and are called active shapes models for their capacity of taking into account synthetically all the shapes of the training set. In the face recognition field, after the shape model of a facial feature has been found, ASM allow us to verify if the face under test has a feature similar to the one of a person stored in the training set. According to Cootes and Taylor [3] the process of the facial shape model creation by ASM consists of the following four steps:

- Landmarking the features of the face images dealing with the subject to be recognized;
- Alignment of the shapes for all the identified features;
- Modeling the shape variability;
- Choice of the number of modes of the shape variation.

To build the model of the shape it is necessary to start from a contour specified by some landmark. This procedure has to be repeated for each shape in the training set. Landmarks selection is a very important step in the model building process because the ASM technique works by modeling how the different landmarks move altogether when the shape varies. As a consequence, any ASM based recognition method will fail if the landmarks do not correctly contour all the shapes in the training set. Figure 4 shows, as an example, a manual landmark contouring the left eye of our sample image.

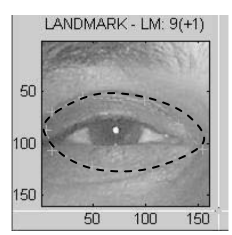

Figure4: Manual landamarking of an eye

In general, landmarks have not to be used during the testing phase, at condition of performing a complex procedure aiming at superimposing and slight deforming the shape model to verify if this feature is possessed by the subject under test. It is easy to understand that the manual landmarking of the features as well as the mentioned recognition procedure is a time consuming activity that may be incompatible with the precision and the real time constraints needed by the recognition phase. For this reason, our tool uses the ASM technique starting from the automated landmarks identified according to the FOA-ACM based technique presented in the previous sections. Moreover the automated landmarking is also used during the testing phase to extract the relevant features of the person under test to be compared with the ones stored in the database.

To allow the ASM technique to be able to compare equivalent landmarks of different shapes in search of a general shape model, these landmarks must be aligned in a common set of reference axes.

The required alignment is obtained by scaling, rotating and translating the training shapes in order to minimize the weighted sum of the squared distances between equivalent points of different shapes. The main steps of the adopted alignment technique, known as Procrustes analysis [7], are illustrated in Figure 5.

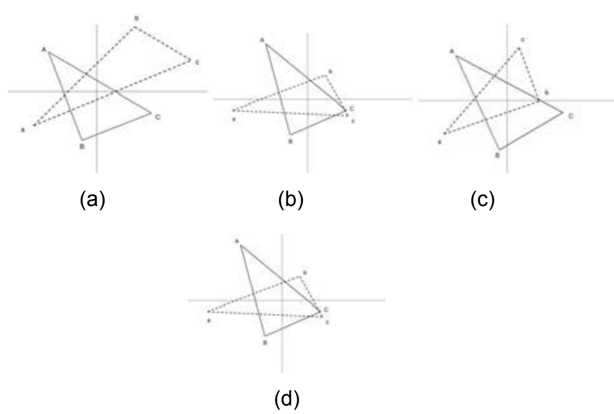

Figure 5: Procrustes analysis: a) initial configuration, b) translation, c) rotation and d) scaling

After having aligned all the shap5es of the training set, each shape can be represented by a vector v_i describing the n points of the i-th shape as follows:

$$v_i = (x_i0, y_i0, x_i1, y_i1, \cdots x_ik, y_ik, \cdots x_in\text{-}1, y_in\text{-}1)^T$$

where (x_ij, y_ij) is the point j of the shape i.

ASM works by examining the statistics of the coordinates of the landmarks in the training set. The final model is a curve having the following form:

$$v = v_{avg} + P * b$$

where v_{avg} is the average shape of the N shapes in the training set, P = (p1,p2, ..., pt) is a set of orthogonal modes of variation given by the first t eigenvectors of the covariance matrix of the N vectors v_i , b = (b1,b2, ..., bt) is a weight vector, one for each eigenvector. Figure 6 shows the model obtained for the left eye (fig.6a) and for the mouth (fig.6b) of the person shown in fig.2b.

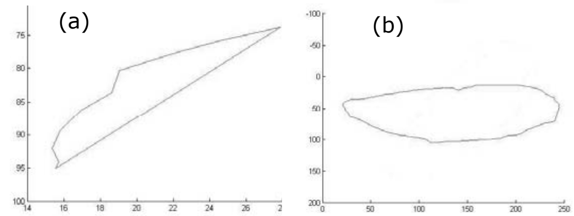

Figure 6: Active shape models for the left eye (a) and for the mouth (b)

V. FACE RECOGNITION METHODOLOGY AND EXPERIMENTAL RESULTS

The methodology proposed in this paper is based on the integration of the techniques described in the previous sections. It develops according to three main phases.

Training phase: ten images with different orientations for each face are considered to build the model for each feature. This procedure is shown in fig.7.

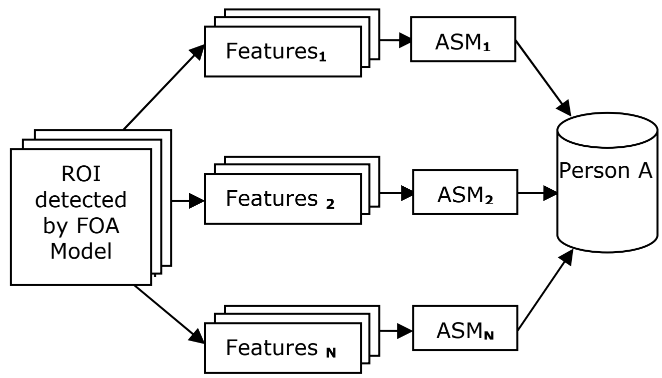

Figure 7: Training procedure for building the features of a person to be recorded in the database

Testing phase: the face recognition is performed as follows:

- Combined analysis of shapes and color variance to locate the region of interest (ROI) of the face [6];
- Visual attention processing to extract the main features of the subject under test;
- Automated initial landmarking of such features and identification of the optimal feature contour by ACM;
- Comparison of the shapes extracted on the basis of the landmarks identified in the previous step with the active shape models contained in the face database by using the Mahalanobis distance.

Face identification: a face is recognized if the matching percentage of the features with the ones recorded in the database is greater than 80%.

The method has been tested by using a database containing 40 models of faces built under three light conditions. The image details are as follows: 400 images dealing with 40 subjects (28 male, 12 female, minimum age 20, maximum age 34). Each subject is characterized by ten images taken for different face expressions, different distances from camera, and different background. The first trial was conducted under optimal light condition, the second one under high luminosity and the third one under low luminosity. Table I shows the number of features that were detected on average, and the percentage of the recognized faces.

Trial	% Detected Features	% Detected Face
1	94%	92%
2	87%	89%
3	74%	82%

Table I Results of the proposed methodology

V. CONCLUDING REMARKS

The proposed method affords the advantage that it is entirely automated thanks to the automated identification of the relevant facial features and related rough landmarks produced during the learning phase by the visual attention module. The optimization of such initial landmarking by ACM allows the ASM technique to build effective shape models starting from a good initial landmarking of the relevant shapes. Better results could be obtained by using models such as the Hidden Markov Models instead of ASM [18]. A discussion on this subject is outside the scope of the paper whose main contribution is to point out how visual attention can be employed to automatically extract the salient features of the faces. Indeed, for frontal images of a face, we have found that the distribution of the FOAs extracted during the testing phase could be sufficient for recognition without resorting to ACM and ASM. This is for further study.

To improve the recognition performances, a feature analysis of a specified target might be employed for top-down biasing the attention module towards the influential features. This enables goal-directed search and better detection of the most significant features of the face. Since the FOA distributions depend on the object features, a similar top-down approach could be applied to address the attention module towards significant objects (e.g., faces) contained in an image. This could allow us to extend the recognition process from a single face in a single picture to locating first and then identifying several people present in a scene, for example taken by a camera.

Testing of the system has shown that it is especially sensitive to contrast variations. This problem can be solved easily by using in the visual attention module property maps opportunely weighted, i.e., with a different influence in the FOA determination. A further improvement might be obtained by including gray level appearance information (in addition to shape) in the model building process. In fact, not only the shape but also gray levels are important for an effective facial feature recognition as illustrated in [2], [3], [4]. This implies that our tool will take into account a second model, called appearance model, to be concatenated to the mentioned shape model, for increasing the recognition performance. Both shape and appearance of the facial features of the person under test will be compared, in the next tool version, with the mentioned concatenated shape and appearance models by using the Mahalanobis distance.

REFERENCES

[1] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Transactions on pattern analysis and machine intelligence*, 15:1,042–1,052, 1993.

[2] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Proc. First British machine Vision Conference*, 245:407–412, 1990.

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape model -theit training and application. *Computer vision graphics and image understanding*, 61:38–59, 1995.

[4] T. F. Cootes, C. J. Taylor, and A. Lanitis. Multi-resolution search using activr shape models. *Proc. on IEEE 12th Int'l Conf. Pattern recognition*, 1:610–612, 1994.

[5] K. Etemad and R. Chellappa. Discriminant analysis fo recognition of human face images. *Proc AVBPA, Lecture Notes in Computer Science*, 99:127–142, 1997.

6] R. C. Gonzales and R. E. Woods. *Digital image processing*. Prentice Hall, New Jersey, 2002.

[7] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40, pages =, 1975.

[8] D. Graham, E. Hadyn, and J. Shepherd. *Perceiving and remembering faces*. New York : Academic Press, London, 1981.

[9] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[10] F. Jiao, S. Li, H. Shum, and D. Schuurmans. Face allignment using statistical models and wavelet features. *Proc. of CVPR'03 IEEE*, pages 1063–1069, 2003.

[11] A. Johnston, H. Hill, and N. Carman. Recognizing faces: Effects of lighting direction, inversion, and brightness reversal. *Perception*, 21:365–375, 1992.

[12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active countorn models. *International Journal of Computer Vision*, pages 321–331, 1998.

[14] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces: probabilistic matching for face recognition. *Proc. on Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 30–35, 1998.

[15] A. Pentlad, B. Moghaddam, T. Starner, O. Oliyide, and M. Turk. View based and modular eigeispaces for face recogntion. *M.I.T. Media Laboratory, Percentual Computing section-Technical report*, 245, 1994.

[16] W. Wang, S. Shan, W. Gao, B. Cao, and Y. Baocai. An improved active shape model for face alignment. *Vision Research*, 40:1489–1506, 2000.

[17] A. L. Yuille. Deformable templates for face detection. *J. Cogn. Neurosci.*, 3:59–70, 1991.

[18] A.V. Nefian and M.H. Hayes. Hidden Markov Models For Face Recognition. *Proc, of the IEEE Conf. on Acoustics, Speech and Signal Processing*, 1998.