FACIAL FEATURE EXTRACTION IN FRONTAL VIEWS USING BIOMETRIC ANALOGIES

Sofia Tsekeridou and Ioannis Pitas
Dept. of Informatics, Aristotle Univ. of Thessaloniki,
Box 451, Thessaloniki 54006, GREECE
Tel: +30 31 996304; fax: +30 31 996304
e-mail: pitas@zeus.csd.auth.gr

ABSTRACT

Face detection and facial feature extraction are considered to be key requirements in many applications, such as access control systems, model-based video coding, content-based video browsing and retrieval. Thus, accurate face localization and facial feature extraction are most desirable. A face detection and facial feature extraction in frontal views algorithm is described in this paper. The algorithm is based on principles described in [1] but extends the work by considering: *(a)* the mirror-symmetry of the face in the vertical direction and *(b)* facial biometric analogies depending on the size of the face estimated by the face localization method. Further improvements have been added to the face localization method to enhance its performance. The proposed algorithm has been applied to frontal views extracted from the European ACTS M2VTS database with very good results.

1 INTRODUCTION

Face localization and facial feature extraction have been active research topics due to their extensive range of applications such as access control systems, model-based video coding, content-based video segmentation, retrieval and browsing, person identification. Due to variations in illumination, background, visual angle or facial expressions, however, face recognition becomes a complex task.

Accurate face localization plays a key role in the successive estimation of facial features. Among other, face localization approaches have focused on the use of shape information (e.g. ellipse-fitting methods) [3], mosaic images [2] or color information [1]. In the latter case, color segmentation is performed keeping only regions with skin-like colors. A usual problem in this approach is the misclassification of irrelevant regions as facial ones due to inaccurate color segmentation. In this paper, the color segmentation approach of [1] is adopted in the face detection mechanism. However, enhancements in the threshold levels of the HSV domain as well as novel design ideas in the final determination of the face object based on biometric face analogies are introduced which tend to increase the flexibility of the face detection module.

A common approach for facial feature extraction is the evaluation of minima and maxima locations in the vertical and horizontal greylevel profiles [1] of the face region determined in the previous step. It is understandable that compensation for head rotation increases the flexibility and robustness of such an approach. A method for correctly evaluating the rotation angle of the head is proposed based on the mirror-symmetry of the face in frontal views. A weighted block matching technique for the estimation of eye locations is adopted which proves to give remarkable results. Further estimation of facial features is performed based on restrictions of their locations according to facial biometric analogies.

The proposed algorithm has been tested on frontal or semi-frontal views extracted from the European ACTS M2VTS database [5]. Frontal views from 37 different persons have been used to evaluate the performance of the proposed method. The simulation results prove the flexibility and robustness of the proposed face localization and facial feature extraction modules in slightly varying lightning conditions, scaling factors or rotation angles.

2 FACE LOCALIZATION

It is well known that accurate face localization enables more robust facial feature extraction. Faces are characterized by their skin-like color, their elliptical shape and the x/y ratio of their surface. The proposed face localization module initially performs color segmentation in a coarse resolution of the initial image to determine an initial map of all pixels having colors close to flesh ones. Refinement of this map follows. The final map is further processed by an object detection algorithm in order to recognize different skin-colored objects in the image. Subsequently, the object among the ones found that is most probably the face is determined. A more detailed description of the face localization algorithm steps is given in the following.

2.1 Color Segmentation

Since skin-like colors can be better characterized by hue and saturation values, color transformation from the RGB to the HSV color space is initially performed [1]. Pixels are afterwards selected having skin-like colors by setting appropriate thresholds to the hue, saturation and value values, specifically $0^0 \leq H \leq 25^0$, $335^0 \leq H \leq 360^0$, $0.2 \leq S \leq 0.6$ and $V \geq 0.4$ (which vary from the ones used in [1]). The hue H range restricts segmentation in reddish colors and the saturation S range ensures exclusion of pure red or very dark red colors and both account for small variations in lighting conditions. The value V thresholding is introduced to discard dark colors (such as dark hair). The simulation results prove the effectiveness of the mentioned thresholds even with small variations in lightning conditions. At this step, an initial map of pixels having skin-like colors is determined. Color segmentation and the next two steps are performed at a coarse resolution of the initial image.

2.2 Initial Map Refinement

After an estimate of the initial map of pixels having skin-like colors has been evaluated, refinement succeeds by discarding secluded assigned ones or by adding non-assigned ones appropriately located in the neighbourhood of assigned ones.

2.3 Object Detection

The next task is to detect different objects in the image. Object determination is performed by an iterative merging algorithm which considers 4-neighbourhood connectivity and which attempts to fill the "holes" that are possibly detected in a compound object (e.g. the regions of eyes that color segmentation has discarded). The merging algorithm starts from the left-most assigned pixel of the currently considered object and proceeds by assigning the next in line pixels to the current object if they are assumed to be connected with previously assigned to the same object pixels. Start and end points are saved and search in the next row is initialized from these points on. The algorithm is performed iteratively until no more connected to the current object pixels are detected. The process is computationally efficient and only a few iterations are required.

2.4 Face Region Detection

After the different skin-colored objects have been extracted, the one that possibly represents the face is detected. The decision is based on the size of the object with respect to the size of the image as well as on the ratio of its width to height which must be in a certain range considering the facial x/y ratio. In cases where only one person appears in the scene such a decision method performs satisfactorily. After the decision is made, all other objects are discarded from the initial map, leaving only the face object. This final map is called face map.

3 FACIAL FEATURE EXTRACTION

After having localised the head, facial feature extraction is pursued. Face detection by its elliptical shape is proposed in [1] and the rotation angle of the head is estimated by the rotation angle of the estimated ellipse. Although in principle this idea is correct, the rotation angle of the estimated ellipse is not always the actual head rotation angle. In the proposed method, the head rotation angle is determined by taking into account the face mirror-symmetry. Initialization is performed by roughly estimating the x- and y-positions of the eyes. Blocks with dimensions relevant to the face width are initialized in the regions of the eyes and a weighted block matching approach estimates their exact location and the head rotation angle. Head rotation compensation is successively performed and feature locations are found by minima analysis to the rotated face profiles. Then, the coordinate system is rotated back to the initial one along with the feature locations. A more detailed description is presented in the following.

3.1 Coarse Eye Locations Estimation

The y-position of the eyes is initially evaluated by minima analysis in the vertical greylevel profile of the face. The latter is derived considering only the pixels of the face map in all three color channels. Eyes are searched for in the upper half of the face according to face biometric analogies. Minima locations are determined using the first and second derivatives of the vertical face profile smoothed beforehand by an averaging one-dimensional filter of size 3. Minima located in a small neighbourhood near other ones with less greylevel value than the former are discarded. The pair of the remaining minima whose distance in pixels lies in a predefined range with respect to the width of the face is selected as the eyebrows and eyes. Thus, an initial estimate of the y-position of the eyes is extracted. In succession, the horizontal eye greylevel profile is evaluated from a region whose height lies in the neighbourhood of the initial eye y-locations. The maximum value of this profile represents the x-center of the face, whereas significant minima from either side of the center whose distances from the it are almost equal, indicate the x-locations of the eyes. Use of the horizontal eye profile proves effective in cases of semi-frontal views, since better estimates of the eye x-positions can be obtained. At the end of this step, a coarse estimate of the eye locations has been calculated.

3.2 Fine Eye Location and Head Rotation Angle Estimation

Once a first estimate of the left and right eye centers (l_x, l_y) and (r_x, r_y) is available, exact calculation of the

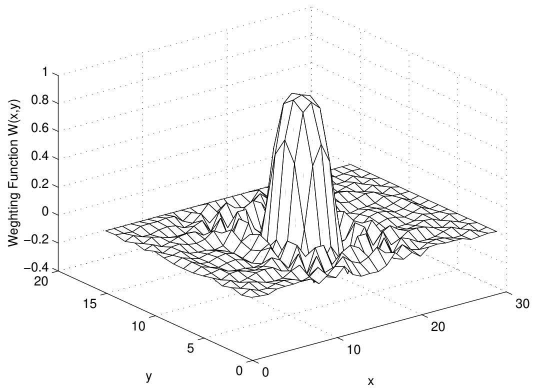

Figure 1: An example of the weighting function $W(x, y)$.

eyes is attempted. In this attempt, the face mirror-symmetry is accounted for [4]. Specifically, a weighted block matching method is used. Initial rectangular eye blocks $b_x \times b_y$, their size being relevant to the face width, are centered around the initially estimated eye positions and search regions of size $((2s_x + 1) \times (2s_y + 1)$ are defined. The centers of the blocks for which the following Matching Error E:

$$
\begin{aligned}
E \;=\; & \sum_{x=0}^{b_x}\sum_{y=0}^{b_y}\sum_{i=1}^{3}[|W(x,y)I_i(l_x - \frac{b_x}{2} + x, l_y - \frac{b_y}{2} + y) \\
& + W(b_x - x, y)I_i(r_x + \frac{b_x}{2} - x, r_y - \frac{b_y}{2} + y)|] \quad (1)
\end{aligned}
$$

is minimized, represent the centers of the eyes. In (1), $I_i(x, y)$, $i = 1, 2, 3$ represent the red, green and blue color channels respectively. $W(x, y)$ is a two-dimensional weighting function of the form:

$$W(x, y) = \frac{sin((x^2 + y^2)/c)}{(x^2 + y^2)/c} \quad (2)$$

an example of which can be seen in Figure 1, that attempts to weight mostly the center of the blocks, which, in the case of eyes, represent the irises that have low brightness and thus keep the total Matching Error low. Furthermore, $W(x, y)$ attempts to negatively weight the horizontal left and right neighbourhoods of the block center, which again in the case of eyes are the white regions of them and thus keep the Matching Error low. The rest parts of the blocks are weighted with almost zero values since they are thought of no significance. In this way, it is proved that very good estimates of the eye centers are extracted, once the initial estimates are not totally wrong. Factor c is analogous to the eye block width.

The head rotation angle θ is estimated by:

$$\theta = \arctan \frac{D_y}{D_x} \quad (3)$$

where D_x and D_y are the horizontal and vertical distances in pixels of the already evaluated eye centers, respectively.

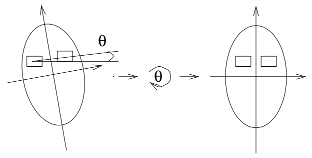

Figure 2: Head rotation compensation by θ.

3.3 Head Rotation Compensation and Feature Extraction

Head rotation compensation is successively performed to rotate the head in the vertical position, as shown in Figure 2. In such a way, better feature detection can be performed. Re-estimation of the vertical face profile is done in the new coordinate system (O, x_r, y_r). Eye positions and the center between the eyes are known (rotated also in (O, x_r, y_r)). Search for minima locations to the vertical face profile is performed using its first and second derivatives, in the way described in 3.1. Eyebrows are searched for in the upper half of the face and the ones lying in a predefined range of distances before the known eye y-positions are selected. Mouth center and chin centers in the y direction are searched for in the lower part of the face. The pair, whose distance lies in a predefined range, is selected. The nose y-position is the position of that minimum between eyes and mouth that is appropriately distanced from both eyes and mouth center. Predefined ranges are set according to facial biometric analogies and depend on the face width. The nose, mouth and chin centers in the x direction are set equal to the x position of the center between the eyes, a true fact for vertically aligned faces, if no strange facial gestures are made. After feature extraction in (O, x_r, y_r) has been accomplished, all feature locations are rotated back to the initial coordinate system (O, x, y), which yield the final feature locations.

4 SIMULATION RESULTS

Frontal or semi-frontal views of 37 different persons from the M2VTS database have been used to evaluate the performance of the proposed method. Results from the face localization algorithm are illustrated in Figure 3. A success rate of correctly detected faces equal to 97.3% is achieved by the face detection method. Face localization fails only in cases where a beard exists. In such cases, only half of the face is detected.

The facial feature extraction method locates candidates for facial features for all the 37 persons of the database. Table 1 summarizes the success rates achieved for correctly located facial features. For comparison purposes, the success rates achieved by the method of [1] are also shown, for detected features. It is evident that better face detection results in better facial feature extraction. Since facial biometric analogies are considered for the estimation of feature locations with respect to the eye locations, wrong estimation of the eyes results

Figure 3: Results obtained by the face localization method (top) and the head rotation compensation method (bottom).

Table 1: Face and Facial feature detection success rates for the M2VTS database.

Feature	Success Rate	Suc. Rate of [1]
Facial Region	97.3%	97.3%
Left Eyebrow	75.67%	67.23%
Right Eyebrow	75.67%	67.23%
Left Eye	81.08%	69.56%
Right Eye	81.08%	69.56%
Nose	83.78%	91.18%
Mouth	86.48%	89.24%
Chin	89.18%	83.42%
All Features	81.85%	76.77%

in possible wrong evaluation of the rest facial features. In the other hand, correct extraction of the eyes depends on the initial estimates. However, once eyes have been successfully extracted, all other features are, in most cases, correctly evaluated. Results of the head rotation compensation method are illustrated in Figure 3. Examples of correct facial feature extraction can be viewed in Figure 4. A very satisfactory estimation of the eye and iris regions is easily observed.

5 CONCLUSIONS

A face localization and facial feature extraction method in frontal views is proposed, extending the work of [1]. It takes under consideration the mirror-symmetry of the face and face biometric analogies. A 81.85% success rate of correctly detected features is achieved. The eye localization approach by a weighting block matching technique gives remarkable results.

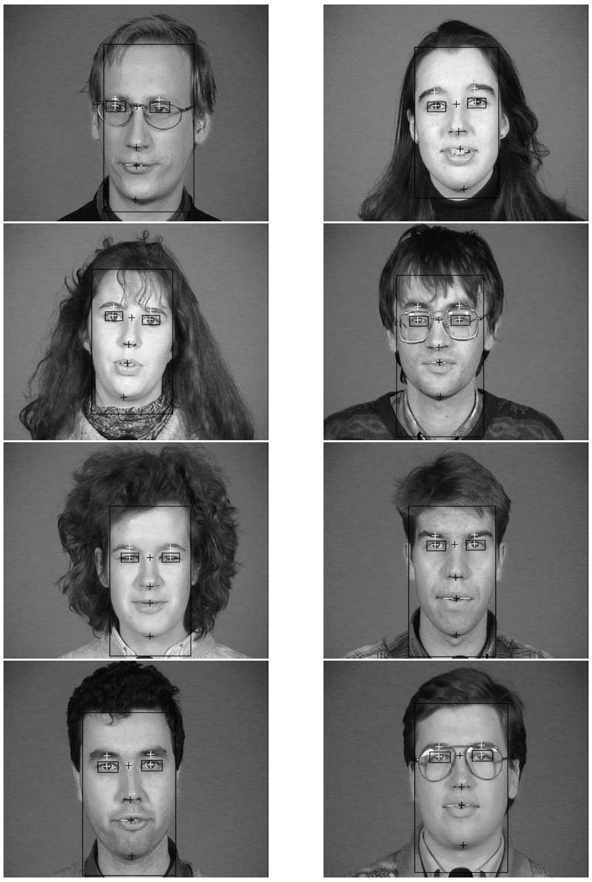

Figure 4: Results obtained by the facial feature extraction method.

References

[1] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking", *Image Communication*, Elsevier, to appear, 1997.

[2] C. Kotropoulos, I. Pitas, S. Fischer and B. Duc, "Face authentication using Morphological Dynamic Link Architecture", *First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Crans-Montana, Switzerland, pp. 169-176, 12-14 March 1997.

[3] A. Eleftheriadis and A. Jacquin, "Automatic face location, detection and tracking for model-assisted coding of video teleconferencing sequences at low bit-rates", *Signal Processing: Image Communication*, vol. 7, no. 3, pp. 231-248, July 1995.

[4] David Beymer, "Face Recognition Under Varying Pose", *in Proc. IEEE Computer Vision and Pattern Recognition*, pp. 756-761, Seattle, USA, June 1994.

[5] S. Pigeon and L. Vandendorpe, "The M2VTS multimodal face database", *in Proc. of the First Int. Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA'97)*, Crans-Montana, Switzerland, 12-14 March 1997.