ELSEVIER

Pattern Recognition Letters 23 (2002) 1623–1629

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Projection based method for segmentation of human face and its evaluation

Selin Baskan [a], M. Mete Bulut [a], Volkan Atalay [b,*]

[a] Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara 06531, Turkey

[b] Department of Computer Engineering, Middle East Technical University, Ankara 06531, Turkey

Received 30 March 2001; received in revised form 12 October 2001

Abstract

We detect facial features and then circumscribe each facial feature with the smallest rectangle possible by using vertical and horizontal gray value projections of pixels. The result is evaluated with respect to the manually located enclosing rectangle on the images of a publicly available database.

Keywords: Facial feature segmentation; Face detection; Gray value projections; Segmentation evaluation

1. Introduction

Considering face as an image pattern, it is a challenge to detect faces and segment into its constituent features, since faces can be very vague, but still have the same basic structure and content. Human face may change its appearance because of facial expression, beard, moustache, hairstyle, make-up, glasses, aging, surgery, etc. In addition to these internal variations, external distortions such as the scale, lighting, position, tilt and orientation of the face must also be considered. Furthermore, it should be noted that a complex background in an image and the high degree of deformability of mouth and eyes make far more difficult to locate faces and facial features.

The goal of our study is to detect facial features such as eyebrows, eyes, nose, mouth and ears, to circumscribe the detected feature with the smallest enclosing rectangle possible and evaluate the result with respect to the manually located enclosing rectangle. We do not extract the shape of the facial features and we assume that such a step follows the result of our work. The contribution of this paper is to describe a complete system in which facial feature segmentation is totally based on projections of pixel gray values and which is tested objectively and quantitatively on a widely accepted face database publicly available. The parameters of the system are tuned on a training set of face images, which is different-in terms of people and capturing conditions—than the test set of face images. In the face detection part of the proposed system, we follow the traditional approach using

* Corresponding author. Tel.: +90-312-2105576; fax: +90-312-2101259.

E-mail address: volkan@ceng.metu.edu.tr (V. Atalay).

0167-8655/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-8655(02)00037-5

color and shape information. Based on the horizontal and vertical projections of pixel gray values, the system circumscribes each facial feature with the smallest rectangle, which is compared with manually extracted rectangle around the corresponding facial feature. We propose two error measures to compare two enclosing rectangles and thus we are able to evaluate quantitatively the facial feature extraction system. The rectangle which circumscribes a facial feature can be used for two purposes: (i) it can be taken as the initial (rough) contour to extract the shape of the facial feature by the use of deformable models or active contour models (Chow and Li, 1993; Kampmann and Zhang, 1998; Yin and Basu, 1999), (ii) the data inside the rectangle can be used to extract features for face recognition or for other purposes (Moghaddam and Pentland, 1997). In our study, we assume that the background is not complex and there is only a single face in an input image. Further, we suppose that the image quality and resolution is sufficient enough, the illumination is uniform and the input images are color images. However, no restrictions on wear, glasses, make-up, hairstyle, beard, etc. are imposed.

There are several proposed solutions to detect faces and facial features (Yuille et al., 1989; Samal and Iyengar, 1992; Yang and Huang, 1994; Chellappa et al., 1995; Forchheimer et al., 1996; Reinders et al., 1996; Rowley et al., 1998; Liu and Wu, 2000). Face detection based on the skin color by filtering is a very popular method (Lee et al., 1996; Saber and Tekalp, 1996; Sobottka and Pitas, 1996). In general, color information is not sufficient alone and therefore ellipse fitting methods are employed (Sobottka and Pitas, 1996) to approximate the shape of the face since human face resembles to an ellipse. For facial feature extraction, one of the very popular methods in geometric-feature based approach is the use of vertical and horizontal projections, which have been proposed and used by several researchers. The projections may be employed after taking the Laplacian (Kanade, 1973) or the first derivative of the image (Brunelli and Poggio, 1993) as well as directly on the intensity values (Sobottka and Pitas, 1996). Projection based methods have been used particularly to find coarsely the position of the facial features which is then followed by a template matching algorithm or another method for refining the result (Chow and Li, 1993; Kampmann and Zhang, 1998; Yin and Basu, 1999). Bernögger et al. (1998) and Yin and Basu (1999) focus on the detection, tracking and modeling of movements of facial features but particularly of eyes. For detection, their general approach is to first find a coarse region in which the facial feature is guaranteed to be and then use a deformable template model to extract the shape. In one of the steps of their method, horizontal projections are used to detect the upper and lower eye lid. In fact this is the part the most related with our work. Kampmann and Zhang (1998) estimate eye, eyebrow and nose features. Eye is estimated by a deformable template. On the other hand, the detection of eyebrow and nose features depends on the face geometry and the pixel gray values. However, they do not present a formal algorithm. Chow and Li (1993) describe a model based system which first localizes face and its features and then refines the locations and shapes of the eye and mouth. For face and eye and mouth localization, valley regions are detected using morphological filtering and component analysis. Vertical projection on which our method is based has not been totally exploited in the previous studies. Although the number of methods and approaches is abundant, no formal way of evaluation of facial feature detection has been reported in the literature. Furthermore, only a few number of studies gives qualitative results about the performance.

This paper is organized as follows. First, we describe how the face is detected in an image and how the input image is enhanced by traditional preprocessing steps. The segmentation of facial features on a preprocessed image is based on the use of horizontal and vertical projections and in Section 3 this procedure is explained. In Section 4, experimental results on a publicly available face database are given.

2. Face detection and preprocessing

The face detection and preprocessing steps include several traditional image processing methods

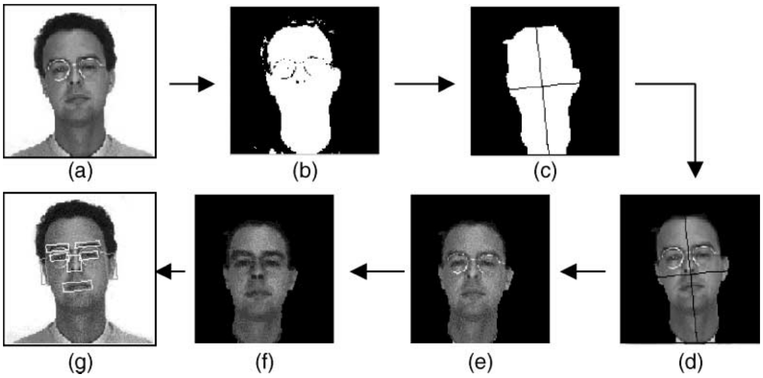

Fig. 1. (a) Original input image, after (b) Skin Filter-1, (c) connected components labeling, closing and ellipse fitting, (d) grayscale image multiplied by the silhouette image, (e) head normalization, (f) grayscale erosion, and (g) the output of facial feature segmentation.

which are applied altogether to obtain a better input data for facial feature segmentation. For face detection, we follow the traditional approach, which is based on color segmentation and shape extraction. Fig. 1 shows the results of the face detection and preprocessing steps on a sample input image. The details of the steps are given as follows.

- *Color segmentation:* Color segmentation algorithm employs skin color filtering. It has been found that the human skin color distribution is limited to a relatively small part of the hue-saturation color space which makes possible to decide with a high certainty, whether a pixel is a skin pixel or not. For this purpose two filters are used.
  - *Skin Filter-1* is designed to extract the skin-colored regions from the image using the following thresholds: $0.23 \leqslant S \leqslant 0.69$, $0° \leqslant H \leqslant 40°$ where S indicates the saturation component and H the hue component of Hue-saturation-intensity representation of color. The result of Skin Filter-1 applied on an input original image given in Fig. 1a is shown in Fig. 1b.
  - For robustness, a second filter, *Skin Filter-2* with the following thresholds: $0.23 \leqslant S \leqslant 0.69$, $0° \leqslant H \leqslant 40°$, $S' > 0.25$, where $S'(x, y)$ corresponds to the saturation value of the pixel (x, y) of the negative image.

  For a face image, both of the filters are applied and the one, which gives the better shape of the face, is selected.

- *Connected components labeling:* It is assumed that the face occupies the greatest part of the image, and hence the connected component with the largest area is the face. To eliminate the components other than the face, connected components labeling algorithm is applied.
- *Morphological closing:* For smoothing purposes, we use morphological closing operator in which the size of the structuring element is made to depend on the size of the face component.
- *Ellipse fitting:* The oval shape of a face can then be approximated by an ellipse. The best-fit ellipse is computed on the base of moments. The application of the connected components labeling, closing and ellipse fitting on the image shown in Fig. 1b is given in Fig. 1c.
- *Image multiplication:* Intensity values of the original image are then multiplied by the pixel values of the silhouette image as shown in Fig. 1d.
- *Normalization:* In the facial feature extraction step, it is assumed that the facial features are horizontally elongated inside the face. If the face is tilted, this assumption is no longer valid, and a normalization of the orientation of the face is necessary. To obtain this normalization, a rotation of the interior of the connected component

is performed. The result of normalization is shown in Fig. 1e.

- *Grayscale erosion:* After normalization, a grayscale erosion is applied to enhance the dark regions (facial features). The result of grayscale erosion is shown in Fig. 1f.

3. Facial feature segmentation based on gray value reliefs

Facial feature segmentation is based on the observation that facial features differ from the rest of the face because of their low brightness (Sobottka and Pitas, 1996). Therefore, facial features are determined by searching for minima in the gray value relief, which is simply the projection of pixel gray values. There are two kinds of reliefs, *X*- and *Y*-Relief that are computed by considering the average of pixel values of the segmented face region along the vertical direction (columns) and along the horizontal direction (rows), respectively. It is assumed that each facial feature generates a minimum in *Y*-Relief and has specific *X*-Relief characteristics. In Fig. 2, the minimum in *Y*-Relief for each facial feature and the corresponding

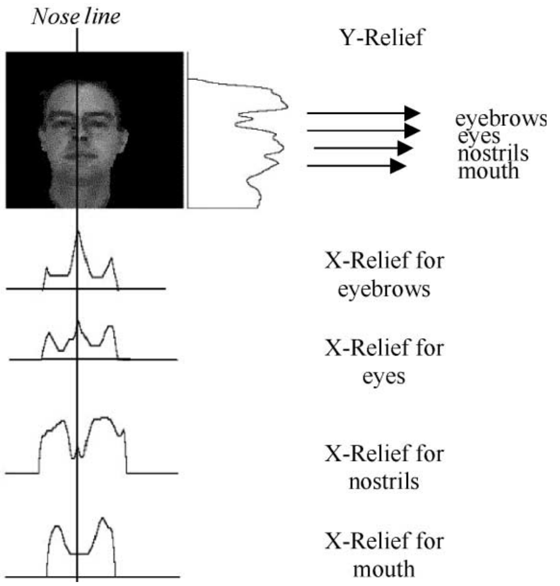

Fig. 2. *Y*- and *X*-Reliefs for a sample image.

X-Reliefs are shown. Remark that nose line is the line that passes through the nose and it acts as the symmetry axis for the face. Assuming that the face in the image is not upside-down, it is expected that the first significant minimum on *Y*-Relief correspond to the eyebrows, the second minimum correspond to the eyes, third to nostrils, fourth to mouth, and the last to chin.

The position of a minimum in *Y*-Relief along with the shape of corresponding *X*-Relief are used to match with a facial feature. In *Y*-Relief, usually the number of minima is greater than the number of features. Therefore, for robustness, in addition to the gray value relief information, the anthropometrical (face geometry) information, which gives cues about the relative positions of facial features with respect to each other, is used as well. In *Y*-Relief, the place of minimum relative to face length and the significance of the minimum are the most important factors to determine the type of the facial feature. After extensive experimentation with a training set of face images, a characteristic (typical) *X*-Relief is derived for each facial feature. The facial feature detection algorithm can be given as follows:

```
for each facial feature, starting from top (eyebrows) going to the bottom (mouth)

{
  repeat
    consider next minimum in Y-Relief
    matched = false
    if the current Y-Relief minimum fits the
    characteristics
      then
      {
        compute upper and lower boundaries
        extract corresponding X-Relief
        compute left and right boundaries
        if extracted X-Relief satisfies the
        facial feature's characteristics
        then matched = true
      }
  until matched or last minimum in Y-Relief
}
```

Y-Relief is used to detect upper and lower boundaries of a facial feature while *X*-Relief de-

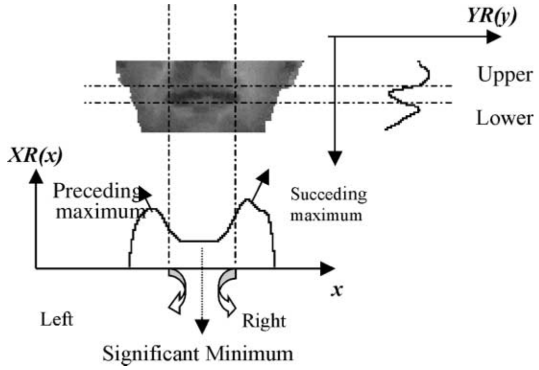

Fig. 3. Y- and X-Relief characteristics of mouth.

termines left and right boundaries. To describe in detail the procedure to compute these boundaries, X- and Y-Reliefs for mouth region of the face presented in Fig. 1f are shown in Fig. 3. In Y-Relief, starting from the significant minimum under consideration and going backwards the preceding maximum, upper boundary is the point (row) at which the highest change occurs. Lower boundary is determined similarly, but between the significant minimum and the succeeding maximum. X-Relief is used to determine the left and right boundaries of a facial feature, by a procedure similar to that explained to find upper and lower boundaries in Y-Relief. These boundaries, in fact determine the enclosing rectangle for a facial feature. Once a pair of X- and Y-Reliefs are extracted, the similarity between the characteristic (typical) reliefs and the extracted candidate reliefs is computed using fuzzy set theory.

Now, we explain in detail the criteria employed for the mouth. Mouth lies in the middle-lower part of the head. Generally, there exists a middle point approximately on a horizontal line in mouth region in which the intensity is lowest; this line is known as the valley line. It is the valley line, which generates a significant minimum in the Y-Relief. If the mouth is closed, this line is located where the lips meet. If the mouth is open, it is positioned along the upper boundary of the lower lip or the lower boundary for the upper lip, or if the mouth is widely open, between the teeth. In any of these cases, the valley line passes through the corners of the mouth. X-Relief calculated along the valley line has the shape of a basin as shown in Fig. 3. The corners of the mouth are detected by the left and right edges of the basin as explained previously in this section. Two X-Relief models are developed for mouth; one for open mouth, the other for closed mouth. For example, when the mouth is closed, there is only one significant minimum in X-Relief and its employed characteristic criteria are given below.

- *Mouth length:* length of the mouth relative to face length lies in a certain range.
- *Similar gray values:* preceding and succeeding maxima have similar gray values.
- *Relative significance of the minimum:* ratio of the local minimum to neighboring maxima is less than a certain value.

Similar requirements are used for eyes, eyebrows and nose detection (Baskan, 2000; Baskan et al., 2001). We use a different method for ear detection. The first step is to reduce the search space using the already available geometrical constraints. Ear region lies somewhere between the eyebrow region and the mouth region. The second step is to extract the contour of the specified region and represent it by a chain code, which enables to perform a template matching like operation (Baskan, 2000).

4. Experimental evaluation and discussion

In order to determine the X-Relief characteristics of each facial feature, we have used an in-house training set of 100 color facial images. For testing purposes, AR face database (Martinez and Benavente, 1998) is used. This face database contains frontal views of 125 individuals with different facial expressions that make a total of 375 color face images. There is no a restriction on beard, moustache, glasses, clothes, make-up, etc. imposed to participants. The developed feature extraction method detects the facial features such as eyebrows, eyes, nose, mouth, and ears and circumscribes each facial feature with the smallest bounding box (rectangle) possible. In order to calculate the error involved in this process, each facial feature in the

Table 1
Rate of missing a facial feature on AR database

	Missing rate (%)
Eyebrows	3.4
Eyes	4.5
Nose	3.7
Mouth	3.4
Ears	16.2

image is marked manually using mouse and the coordinates of these markings are used as ground truth data. Table 1 presents the rate of missing a facial feature in the sense that the facial feature is detected manually but not by the proposed method.

When a facial feature is detected both manually and by the method, the following two criteria are used to compute fitness of two given enclosing rectangles: E_1, distance between the area centers and E_2, ratio of the areas. These errors can be explained better in cooperation with the diagram shown in Fig. 4. Let R_1 be the rectangle drawn manually to circumscribe a facial feature and R_2 be the bounding rectangle calculated by the facial feature extraction method. The center, width and height of R_1 are (x_1, y_1), w_1 and h_1, respectively. Similarly, the center, width and height of R_2 are (x_2, y_2), w_2 and h_2 respectively. The deviation in the center of area is given by $\mathrm{d}x = |x_1 - x_2|$ and $\mathrm{d}y = |y_1 - y_2|$. E_1 is calculated using the Euclidian distance between the center of areas of the two rectangles $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}/w_1h_1$. Let ratio r of the areas of two rectangles be defined as follows:

$$r = \begin{cases} \dfrac{w_1h_1}{w_2h_2}, & \text{if } w_1h_1 > w_2h_2 \\ \dfrac{w_2h_2}{w_1h_1}, & \text{otherwise} \end{cases}$$

A second error, E_2 is defined to be related to r. E_2 is calculated using a function which increases linearly when the ratio r is less than 4 and saturates thereafter. Table 2 gives the mean, μ_1 and standard deviation, σ_1 of E_1 and μ_2 and σ_2 of E_2 for each facial feature of the images in the database. It is easily observed from the error values that E_1 is much smaller than E_2. E_1 gives an idea about the reliability of the feature detection algorithm. The database used for testing purposes includes many people with beards, moustaches, fringes, glasses, and earrings etc., which complicate the feature extraction task considerably. It is also observed that, the bounding rectangle of mouth may be larger than required when the person has beards or moustaches. The same is true for the bounding rectangle of eyes or eyebrows when the person has eyeglasses, etc. These are the possible reasons for larger values of E_2. It can be observed from the results that the proposed method is efficient in extracting the facial features.

The features except ears are detected with a high degree of success. However the ear detection is much more difficult, since in the frontal images only a small part of the ear may be observable, or the hair region may hide some part of the ear.

Some example images along with the segmented features are shown in Fig. 5. As observed from these example images, eyeglasses usually do not create problem in the algorithm. Effect of eyeglasses may be due to the created reflection. Beards and moustaches also create complex problems for

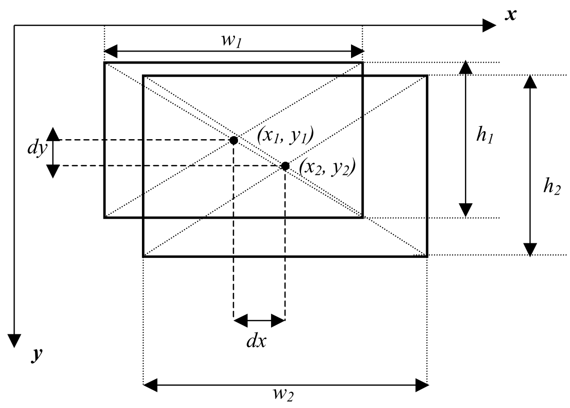

Fig. 4. Parameters used for the comparison of two rectangles.

Table 2
Error statistics if the feature is detected on AR database

	μ_1 (%)	σ_1 (%)	μ_2 (%)	σ_2 (%)
Eyebrows	2.9	0.043	9.7	2.0
Eyes	2.7	0.036	15.5	2.3
Nose	0.6	0.002	16.2	1.2
Mouth	2.5	0.048	10.5	2.6
Ears	12.5	1.252	23.3	9.4

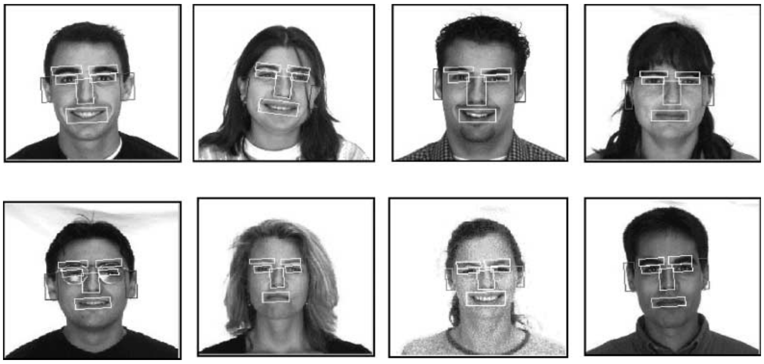

Fig. 5. Sample results from AR database.

feature segmentation. The proposed method is implemented in Java programming language that enables platform independence and it takes approximately 5 s on a Pentium II 233 MHz processor for images of 200×192 pixels. From the experimental evaluation, it is observed that this feature segmentation algorithm is also capable of handling these complex situations up to a certain extent.

References

Baskan, S., 2000. Segmentation of human face. MSc. Thesis, Middle East Technical University, Dept. of Electrical and Electronics Engineering.

Baskan, S., Bulut, M.M., Atalay, V., 2001. Segmentation of human face using gradient based approach. Proceedings of SPIE 4301, 52–63.

Bernögger, S., Yin, L., Basu, A., Pinz, A., 1998. Eye tracking and animation for MPEG-4 coding. In: 14th International Conference On Pattern Recognition, pp. 1281–1284.

Brunelli, R., Poggio, T., 1993. Face recognition: features versus templates. IEEE Trans. Pattern Anal. Machine Intell. 15, 1042–1052.

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

Chow, G., Li, X., 1993. Towards a system for automatic facial feature detection. Pattern Recognition 26, 1739–1755.

Forchheimer, R., Mu, F., Li, H., 1996. Automatic extraction of human facial features. Signal Process. Imag. Comm. 8, 309–332.

Kampmann, M., Zhang, L., 1998. Estimation of eye, eyebrow and nose features in videophone sequences. Int. Workshop on Very Low Bitrate Video Coding, 101–104.

Kanade, T., 1973. Picture Processing by Computer Complex and Recognition of Human Faces. Tech. Report, Kyoto University, Dept. of Inform. Science.

Lee, C.H., Kim, J.S., Park, K.H., 1996. Automatic face location in a complex background using motion and color information. Pattern Recognition 29, 1877–1889.

Liu, A., Wu, F., 2000. Summarization of face objects in a video sequence, http://www.ctr.columbia.edu/~ifongwu/visrep.html.

Martinez, M., Benavente, R., 1998. The AR Face Database. CVC Technical Report #24.

Moghaddam, B., Pentland, A., 1997. Probabilistic visual learning for object representation. IEEE Trans. Pattern Anal. Machine Intell. 19, 696–710.

Reinders, M.J.T., Koch, R.W.C., Gerbands, J.J., 1996. Tracking facial features in image sequences using neural networks. In: Proceedings of 2nd International Conference on Automatic Face and Gesture Recognition, pp. 230–235.

Rowley, H.A., Baluja, S., Kanade, T., 1998. Neural network-based face detection. IEEE Trans. Pattern Anal. Machine Intell. 20, 23–38.

Saber, E., Tekalp, M., 1996. Face detection and facial feature extraction using color shape and symmetry-based cost functions. In: International 13th Conference On Pattern Recognition, pp. 654–657.

Samal, A., Iyengar, P.A., 1992. Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognition 25, 65–77.

Sobottka, K., Pitas, I., 1996. Extraction of facial regions and features using color and shape information. In: 13th International Conference on Pattern Recognition, pp. 421–425.

Yang, G.Z., Huang, T.S., 1994. Human face detection in complex background. Pattern Recognition 27, 53–63.

Yin, L., Basu, A., 1999. Integrating active face tracking with model based coding. Pattern Recognition Lett. 26, 651–657.

Yuille, A.L., Cohen, D.S., Hallinan, P.W., 1989. Feature extraction from faces using deformable templates. Internat. J. Comput. Vision 8, 99–111.