FACE LOCALIZATION AND FACIAL FEATURE EXTRACTION BASED ON SHAPE AND COLOR INFORMATION

Karin Sobottka and Ioannis Pitas

Department of Informatics
University of Thessaloniki 540 06, Greece
E-mail: {sobottka, pitas}@zeus.csd.auth.gr

ABSTRACT

Recognition of human faces out of still images or image sequences is a research field of fast increasing interest. At first, facial regions and facial features like eyes and mouth have to be extracted. In the present paper we propose an approach that copes with problems of these first two steps. We perform face localization based on the observation that human faces are characterized by their oval shape and skin-color, also in the case of varying light conditions. For that we segment faces by evaluating shape and color (HSV) information. Then face hypotheses are verified by searching for facial features inside of the face-like regions. This is done by applying morphological operations and minima localization to intensity images.

Keywords: Face recognition, color image processing.

1. INTRODUCTION

In general the procedure of machine face recognition can be described as follows (Figure 1): Still images or an image sequence and a database of faces are available as input. In a first step, facial regions are segmented. Then facial features are extracted. Afterwards an identification is done by matching the extracted features with features of the database. As result an identification of one or more persons is obtained. Obviously, the order of the first two steps can be interchanged. For example in [1], facial features are extracted first. Then constellations are formed from the pool of candidate feature locations and the most face-like constellation is determined. Also in [2] an approach is presented which derives locations of whole faces from facial parts.

We have decided to segment first the face regions and then to extract facial features. In the case of color images, this processing order allows a very robust analysis, because faces differ significantly from the background by their color and shape. An overview about existing literature on human and machine recognition of faces is given in [3] and [4].

This work was partially supported by the European ACTS-M2VTS project.

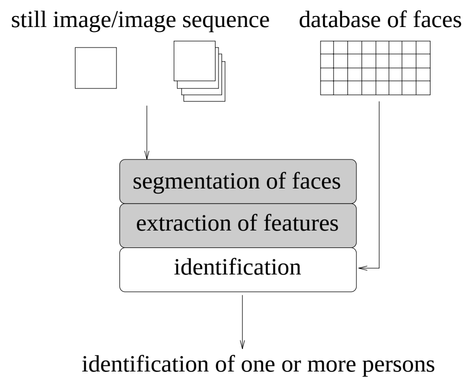


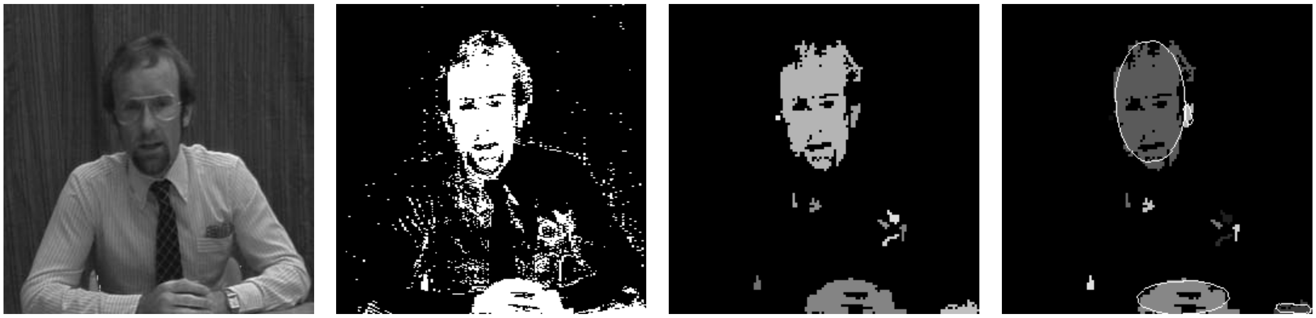
Figure 1: Procedure of machine face recognition

2. FACE LOCALIZATION

Many different approaches for face localization are published in the literature using texture [5], depth [6], shape [7] and color information [8] or combinations of them. However, the detection of facial regions out of scenes with complex background is still a problem.

2.1. Color segmentation

In our approach we first locate skin-like regions by performing color segmentation. As interesting color space we consider the Hue-Saturation-Value (HSV) color space, because it is compatible to the human color perception. Alternatively, similar color spaces (e.g. HSI, HLS) can be used as well. The HSV color space has hexcone shape as illustrated in Figure 3a. Hue (H) is repesented as angle. The purity of colors is defined by the saturation (S), which varies from 0 to 1. The darkness of a color is specified by the value component (V),

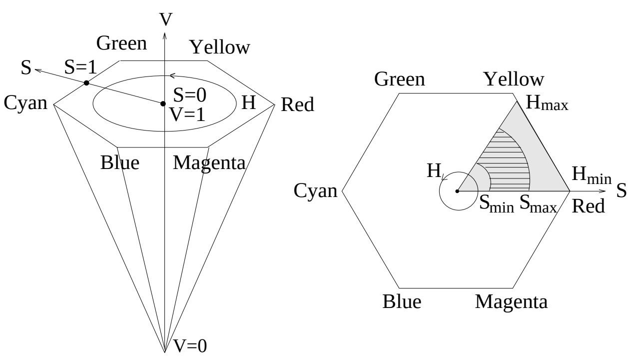


(a) (b) (c) (d)

Figure 2: Detection of facial regions: (a) original image, (b) color segmented image, (c) connected components and (d) best-fit ellipses

which varies also from 0 (root) to 1 (top level).

For the segmentation of skin-like regions it is sufficient to consider hue and saturation as discriminating color information. The hue and saturation domains, which describes the human skin color, can be defined or estimated a priori and used subsequently as reference for any skin color. In our case we have chosen the parameters as follows: $S_{min} = 0.23$, $S_{max} = 0.68$, $H_{min} = 0^0$ and $H_{max} = 50^0$. As it is shown in Figure 3b, this is equivalent to cut a sector out of the hexagon. An example of such a segmentation is shown in Figure 2a,b.



(a) (b)

Figure 3: (a) Hue-Saturation-Value color space (HSV) and (b) skin segmentation in HS space

2.2. Evaluation of shape information

The oval shape of a face can be approximated by an ellipse. Therefore looking for faces in an image means to detect objects with nearly elliptical shape. This can be done based on edges [7] or, as we will show here, based on regions. The advantage of considering regions is that they are more robust against noise and changes in illumination. In a first step connected components are determined by applying a region growing algorithm at a coarse resolution of the segmented image (Figure 2c). Then, for each connected component C with a given minimum size, the best-fit ellipse E is computed on the basis of moments.

An ellipse is exactly defined by its center $(\overline{x}, \overline{y})$, its orientation θ and the length a and b of its minor and major axis. The center $(\overline{x}, \overline{y})$ of the ellipse is given by the center of gravity of the connected component. The orientation θ of the ellipse can be calculated by determining the least moment of inertia:

$$\theta = \frac{1}{2} \cdot arctan \left(\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right) \tag{1}$$

where $\mu_{i,j}$ denotes the central moments of the connected component. The length of major and minor axis of the best-fit ellipse can also be computed by evaluating the moments of inertia. With I_{min} the least and I_{max} the greatest moment of inertia of an ellipse,

$$I_{min} = \sum_{(x,y) \in C} [(x - \overline{x}) cos\theta - (y - \overline{y}) sin\theta]^2 \tag{2}$$

$$I_{max} = \sum_{(x,y) \in C} [(x - \overline{x}) sin\theta - (y - \overline{y}) cos\theta]^2 \tag{3}$$

the length a of the major axis and the length b of the minor axis results in:

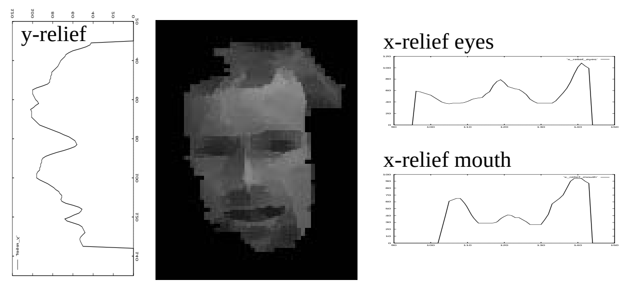
$$a = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{max})^3}{I_{min}}\right]^{1/8} \qquad b = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{min})^3}{I_{max}}\right]^{1/8} \tag{4}$$

To assess, how well the connected component is approximated by its best-fit ellipse, we determine the distance between the connected component and the best-fit ellipse by counting the "holes" inside of the ellipse and the points of the connected component that are outside of the ellipse. The ellipses that are good approximations of connected components are selected and considered as face candidates. In the case of the example we obtain the results shown in Figure 2d. Subsequently the determined face candidates are verified by searching for facial features inside of the connected components.

3. EXTRACTION OF FACIAL FEATURES

Our approach to the extraction of facial features is based on the observation that, in intensity images, eyes and mouth differ from the rest of the face because of their lower brightness. The reason for that is the color of the pupils, the sunken eye-sockets and the light red color of the lips. Even if the eyes are closed, the darkness of the eye sockets is sufficient to extract the eye regions. Therefore, in the following, we consider the intensity information in the interior of the connected component (Figure 7a).

In a preprocessing step, we enhance dark regions by using morphological operations. First we apply a greyscale erosion. Then we improve the contrast of the connected component by an extremum sharpening operation. Results of this preprocessing step are shown in Figure 7b.

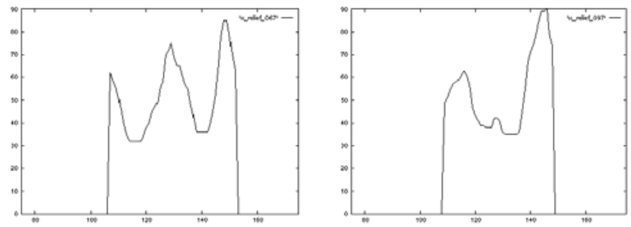


(a) (b) (c)

Figure 4: (a) y-projection, (b) preprocessed face candidate and (c) x-projections

The position of eyes and mouth are determined by evaluating the minima of the topographic greylevel relief. This can be done by using watersheds [9], or as we shall show here, by directly evaluating the x- and y-projections of the greylevel relief (Figure 4). First we compute the mean greylevel of every row of the connected component and we determine minima in the resulting y-relief (Figure 4a). Then for each significant minimum of the y-relief, x-reliefs are computed (Figure 4c). After smoothing the x-reliefs, minima and maxima are detected. Beginning with the uppermost minima in y-direction, we search for minima in x-direction. The positions of the eyes are found, if two minima in x-direction are detected that meet the requirements for eyes concerning relative position inside of head, similarity of greylevel values, significance of maximum between them and ratio of eye distance to head width. A typical example for an x-relief of the eye region is shown in Figure 5a.

The assessment of how well a pair of minima meets the requirements, is done on the basis of fuzzy set the-



(a) (b)

Figure 5: Examples for x-reliefs of (a) eyes and (b) mouth

ory. Thus, we define a membership function for each of these requirements. For example, we use the membership function that is illustrated in Figure 6 to assess the ratio of eye distance to head width.

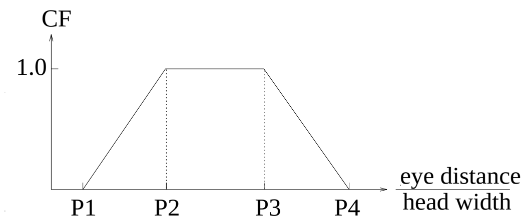
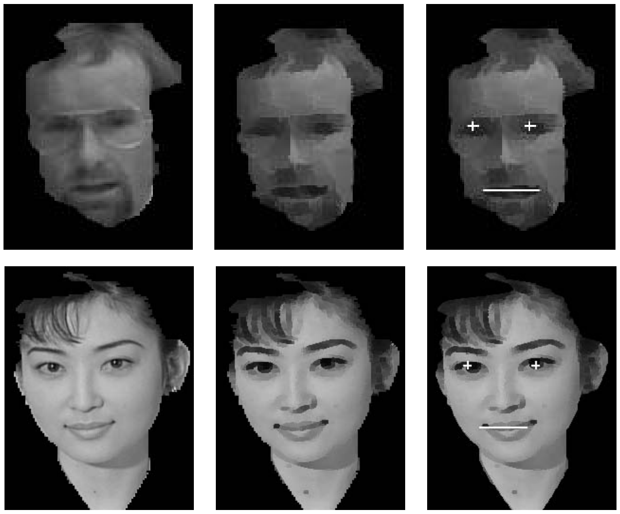


Figure 6: Assessment of the ratio of eye distance to head width

The parameters $P1$, $P2$, $P3$ and $P4$ are defined in dependence on the width of the connected component that is assumed to correspond with the width of the head. On the basis of assessments for these criteria, eye candidates are selected. They have to meet a minimum assessment for each criterion as well as a minimum assessment for the weighted sum of all assessments.

For the mouth a full search through all determined x-reliefs is done as well. As you can see in Figure 5b, the mouth is characterized in x-reliefs by a significant basin with one or two minima inside. For that, mouth candidates are found by looking for two significant maxima that build the borders of the mouth and meet the requirements of the mouth concerning relative head position, significance of minima between them and ratio of mouth width to head width. In analogy to the search for eye candidates, every pair of maxima is assessed, on how well it meets the requirements for the mouth.

After the determination of eyes and mouth candidates, the best constellation of facial features is selected. For that we cluster the candidates for eyes and mouth according to their x-coordinates of the left and right eye or left and right corner of the mouth. Based on the number of votes for a cluster center, the ratio of vertical eye-mouth distance to head height and the horizontally overlapping of mouth and eyes, we choose the best constellation. In the case that only eyes or mouth



(a) (b) (c)

Figure 7: Facial feature extraction: (a) intensity information (b) preprocessed face candidates, (c) detected facial features

candidates were detected, we select the best candidate of them. Results of the detection of facial features are shown in Figure 7c. The eyes and the mouth are well localized in both examples.

4. EVALUATION

To assess the robustness of our approach, we applied our method to an image sequence consisting of 150 frames. An example frame is shown in Figure 2a. Because the face contains features like beard, glasses and changes in facial expressions, the sequence is qualified for testing the algorithm. The results of our evaluation are illustrated in Table 1.

	[%]
facial features	86
eyes	96
mouth	87
correct eyes	98
correct mouth	86

Table 1: Detection rates for facial features

Facial features, that means eyes or mouth, are detected in 86% of the frames.Out of this 86%, in 96% of the frames the eyes are extracted, but only in 98% of these cases correctly. The mouth is detected in 87% of the cases, but only in 86% correctly. These results are very satisfactory.

5. CONCLUSION

In this paper we have presented an approach for the detection of facial regions and features in color images. Facial regions are detected on the basis of color and shape information since faces are significantly characterized by their skin color and oval shape. The extraction of facial features as eyes and mouth is done by evaluating the topographic greylevel relief. Results were shown for two example scenes.

6. REFERENCES

[1] M.C. Burl, T.K. Leung, and P. Perona. Face localization via shape statistics. In *International Workshop on Automatic Face and Gesture Recognition*, pages 154–159, Zurich, Switzerland, June 26-28 1995.

[2] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto. Locating faces and facial parts. In *International Workshop on Automatic Face and Gesture Recognition*, pages 41–46, Zurich, Switzerland, June 26-28 1995.

[3] R. Chellappa, C.L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, May 1995.

[4] M. Bichsel, editor. *International Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, June 26-28 1995. IEEE Computer Society, Swiss Informaticians Society et al., MultiMedia Laboratory, Department of Computer Science, University of Zurich.

[5] Y. Dai and Y. Nakano. Extraction of facial images from complex background using color information and SGLD matrices. In *International Workshop on Automatic Face and Gesture Recognition*, pages 238–242, Zurich, Switzerland, June 26-28 1995.

[6] G. Galicia and A. Zakhor. Depth recovery of human facial features from video sequences. In *IEEE International Conference on Image Processing*, pages 603–606, Washington D.C., USA, October 23-26 1995. IEEE Computer Society Press, Los Alamitos, California.

[7] 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*, 7(3):231–248, Jul 1995.

[8] H. Wu, Q. Chen, and M. Yachida. An application of fuzzy theory: Face detection. In *International Workshop on Automatic Face and Gesture Recognition*, pages 314–319, Zurich, Switzerland, June 26-28 1995.

[9] K. Sobottka and I. Pitas. Localization of facial regions and features in color images. In *4th Open Russian-German Workshop: Pattern Recognition and Image Analysis*, pages 134–138, Valday, The Russian Federation, March 3-9 1996.