Extraction of Facial Regions and Features using Color and Shape Information

Karin Sobottka Ioannis Pitas

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

Abstract

There are many applications for systems coping with the problem of face localization and recognition, e.g. model-based video coding, security systems, mug shot matching. Due to variations in illumination, background, visual angle and facial expressions, the problem of machine face recognition is complex. In this paper we present a robust approach for the extraction of facial regions and features out of color images. First face candidates are located based on color and shape information. Then the topographic greylevel relief of facial regions is evaluated to determine the position of facial features as eyes and mouth. Results are shown for two example scenes.

1. Introduction

Machine face recognition is a research field of fast increasing interest. Although a lot of work has already been done, a robust extraction of facial regions and features out of complex scenes is still a problem. An overview about existing literature on human and machine recognition of faces is given in [2] and [3].

In the first step of face recognition, the localization of facial regions and the detection of facial features, e.g. eyes and mouth, is necessary. In this paper we present an approach that copes with the problems of this first step. As it is described in [1], it is part of a multimodal system for person verification using speech and image information. Our approach for face localization is based on the observation that human faces are characterized by their oval shape and skin-color, also in the case of varying light conditions. Therefore, we locate face-like regions on the base of shape and color information. We employ the Hue-Saturation-Value (HSV) color space by using the RGB to HSV transformation. Alternatively, similar color spaces (e.g. HSI, HLS) can be used as well. The hypotheses for faces are verified by searching for facial features inside the facial regions. We extract facial features based on the observation that eyes and mouth differ from the rest of the face in intensity because of their lower brightness. Thus, dark regions are enhanced in a preprocessing step by applying morphological operations. Afterwards, facial features are located by evaluating the minima and maxima of the topographic greylevel relief.

2. Segmentation of face-like regions

In the field of face localization, several approaches have been published using texture [4], shape [5], depth [6] and color information [11] or combinations of them. However, the detection of facial regions out of scenes with cluttered background is still a problem.

Faces are significantly characterized by their oval shape and specific skin-color. For that reason, we have chosen an approach based on these characteristics. The effectiveness of the use of color information was also shown in [4]. We consider hue and saturation as discriminating color information. These attributes are strongly related to the human perception of color. To obtain robustness against changes in light conditions, we perform region segmentation by considering appropriately defined domains in hue and saturation domain. These domains can be 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 $H_{max} = 50^0$. An example of such a segmentation is shown in Figure 1a,b.

The oval shape of a face can be approximated by an ellipse. Therefore, looking for faces in an image can be performed by detecting objects with elliptical shape. This can be done based on edges [5] or, as we will show here, based on regions. The advantage of consider-

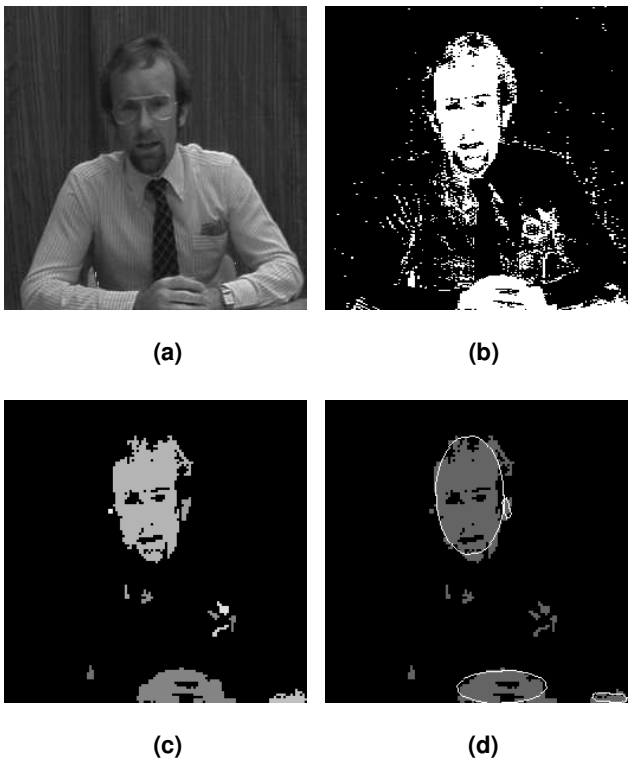


Figure 1. Face segmenation: (a) original color image, (b) skin segmentation, (c) connected components, (d) best-fit ellipses

ing regions is that they are more robust against noise and changes in illumination. In a first step, connected components are determined. Then, we check for each connected component, if its shape is nearly elliptical or not. We determine connected components by applying a region growing algorithm at a coarse resolution of the segmented image. For the image of Figure 1a, we obtain the result shown in Figure 1c. Then, for each connected component C with a given minimum size, the best-fit ellipse E is computed on the base of moments [7]. 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 (Figure 2a). 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 computed by determining the least moment of inertia:

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

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 with orientation θ,

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

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

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

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

On the basis of the already computed elliptical parameters, a reduction of the number of potential face candidates is possible. This is done by applying to each ellipse decision criteria concerning its orientation and the relationship between major and minor axis. For example the orientation has to be in the interval of $[-45^0, +45^0]$. For the remaining candidates we assess, how well the connected component is approximated by its best-fit ellipse. For that purpose the following measure V is evaluated:

$$V = \frac{\sum_{(x,y)\in E}(1 - b(x,y)) + \sum_{(x,y)\in C\setminus E} b(x,y)}{\sum_{(x,y)\in E} 1} \quad (5)$$

with

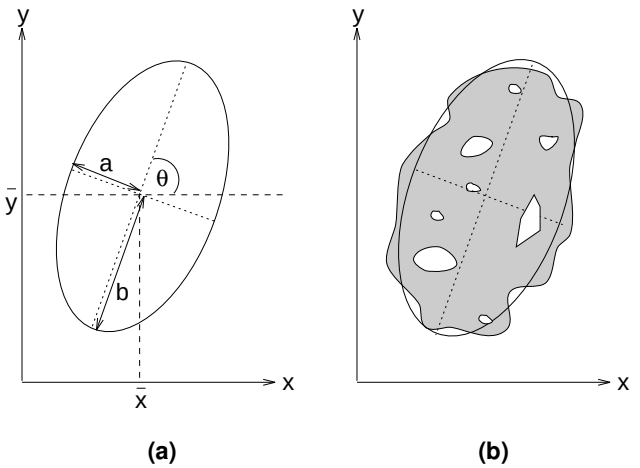
$$b(x,y) = \begin{cases} 1 & \text{if } (x,y) \in C \\ 0 & \text{otherwise} \end{cases}$$

V determines 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 the ellipse (Figure 2b). The ratio of the number of false points to the number of points of the interior of the ellipse is calculated. Based on a threshold on this ratio, 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 Fig. 1d.

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



(a) (b)

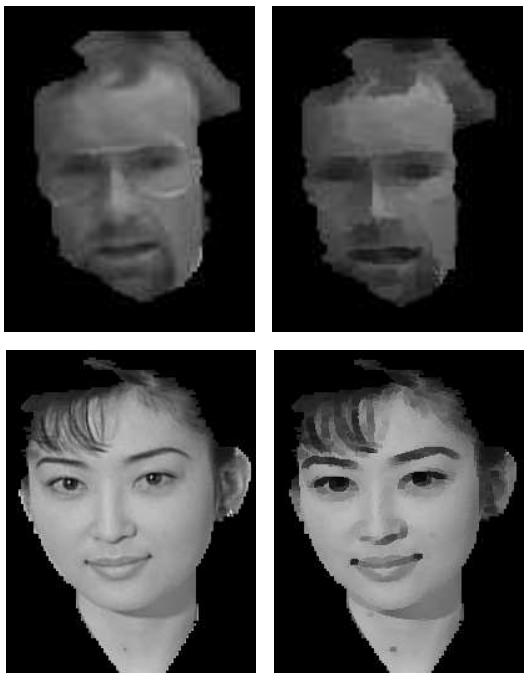
Figure 2. (a) Parameter of an ellipse and (b) Approximation of a connected component

regions. Therefore, in the following, we consider the intensity information in the interior of the connected components (Fig. 3a). In a preprocessing step, we enhance dark regions by using morphological operations. First we apply a greyscale erosion. As structuring element we use a 5×3 rectangle that is elongated in horizontal direction since both, eyes and mouth, are elongated horizontally. Then we improve the contrast of the connected component by the following extremum sharpening operation [8]:

$$g(x,y) = \begin{cases} min & \text{if } f(x,y) - min < max - f(x,y) \\ max & \text{otherwise} \end{cases} \tag{6}$$

Here min and max denotes the minimum and maximum value in the neighbourhood of $f(x,y)$. As neighbourhood we chose again the 5×3 rectangle. Results of this preprocessing step are illustrated in Figure 3b. Eyes and mouth and parts of the hair and beard are emphasized.

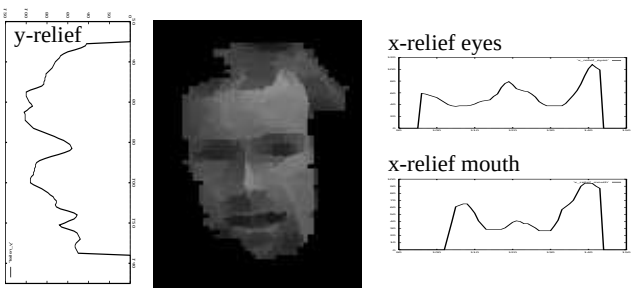
The positions of eyes and mouth are determined by searching for minima in the topographic greylevel relief. This can be done by using watersheds [9], or as we will show here, by directly evaluating the x- and y-projections of the greylevel relief (Figure 4). Since eyes and mouth are horizontally orientated, a normalization of the orientation of the face candidate is necessary. The advantage of this normalization step is that, afterwards, we can search easily for eyes and mouth along the horizontal direction. After normalization, we compute the mean greylevel of every row of the connected component and then determine the minima in the resulting y-relief (Figure 4a). Beginning with the uppermost minima in y-direction, we search for minima in x-direction. The positions of eye candidates are



(a) (b)

Figure 3. (a) Face regions and (b) enhanced facial features

found, if two significant minima in x-direction are detected that meet the requirements for eyes concerning e.g. similarity of greylevel values, ratio of eye distance to head width, relative position inside of head, significance of maximum between them (Figure 4c). The assessment of how well a pair of minima meets the requirements, is done on the basis of fuzzy set theory. Thus, we define a membership function for each of these requirements. For example the membership function for the assessment of the ratio of eye distance to head width is shown in Figure 5a. The parameters $P1$, $P2$, $P3$ and $P4$ are defined in dependence on the width of the connected component that is assumed to



(a) (b) (c)

Figure 4. Detection of facial features

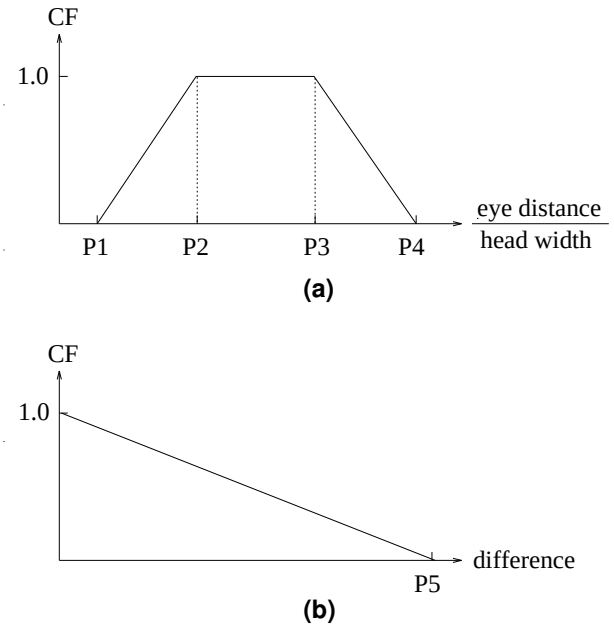


Figure 5. Assessment of criteria: (a) membership function for the ratio of eye distance to head width, (b) membership function for the similarity of greylevel values

correspond to the width of the head. The certainty factor (CF) has the value 1.0, if the measured ratio r is in the range $P2 < r < P3$. If $P1 < r < P2$ or $P3 < r < P4$ holds, the requirement is only partly fulfilled and $0.0 < CF < 1.0$. Otherwise $CF = 0.0$. Another example is the membership function for the assessment of similarity of greylevel values, illustrated in Figure 5b. With increasing greylevel difference of the left and right eye candidate, CF decreases. On the base of these assessments, eye candidates are selected. They have to meet a minimum assessment for each criterium as well as a minimum assessment for the weighted sum of all assessments. Examples of selected eyes candidates, represented as crosses, are illustrated in Figure 6.

For the mouth a full search through all determined x-reliefs is done as well. 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 e.g. ratio of mouth width to head width, relative position inside of head, significance of minimum between them (Figure 4c). In analogy to the search for eye candidates, every pair of maxima is assessed and mouth candidates are selected. Examples of detected mouth candidates, represented as horizontal line segments, are shown in Figure 6.

After eyes and mouth candidates are determined,

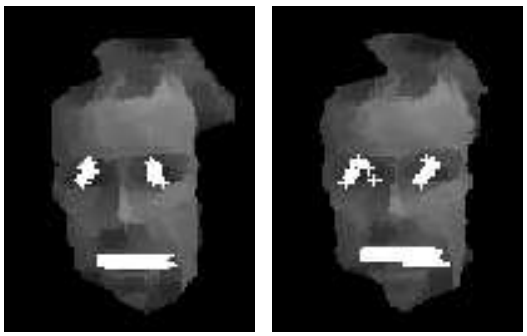


Figure 6. Eyes and mouth candidates

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 overlapping of mouth and eyes in horizontal direction, we choose the best constellation. In the case when only eyes or mouth candidates were detected, we select the best candidate of them. Results of the detection of facial features are shown in Fig. 7. The eyes and the mouth are well localized in both examples.

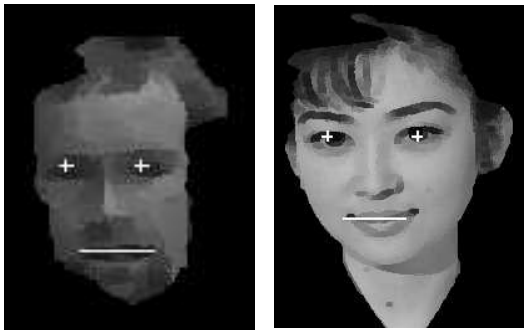


Figure 7. Results of facial feature extraction

4. Evaluation

To assess the robustness of our approach, we applied our method to an image sequence consisting of 150 frames. Examples of frames are shown in Figure 8. Because the face contains features as 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, 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.

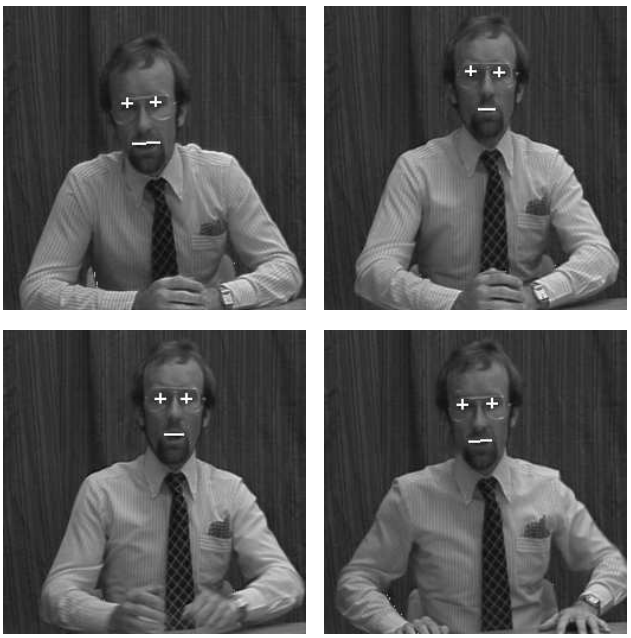


Figure 8. Results on different frames

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

Table 1. Detection rates for facial features

These results are very satisfactory.

5. Conclusion

We have presented an approach that detects facial regions in color images and calculates facial features. Facial parts are determined on the basis of color and shape information. Therefore, we first extract regions with skin-like hue and saturation values and then compute the best-fit ellipse for each of the regions. On the base of the observation that eyes and mouth differ from the rest of the face because of their lower brightness, we first enhance the dark regions inside of the ellipse by applying morphological operations. Then we determine the position of facial features by evaluating the topographic greylevel relief. The success of this method was verified on a large number of color images containing faces.

Acknowledgement

This work has been carried out within the framework of the MAVI project. The financial support is gratefully acknowledged.

References

[1] M. Acheroy, C. Beumier, J. Bigün, G. Chollet, B. Duc, S. Fischer, D. Genoud, P. Lockwood, G. Maitre, S. Pigeon, I. Pitas, K. Sobottka, and L. Vandendorpe. Multi-modal person verification tools using speech and images. In *European Conference on Multimedia Applications, Services and Techniques*, Louvain-La-Neuve, Belgium, May 28-30 1996.

[2] 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.

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

[4] 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.

[5] 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.

[6] G. Galicia and A. Zakhor. Depth recovery of human facial features from video sequences. In *IEEE International Conference on Image Processing*, volume 2, pages 603–606, Washington D.C., USA, October 23-26 1995. IEEE Computer Society Press, Los Alamitos, California.

[7] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.

[8] H. Niemann. *Pattern Analysis and Understanding*. Springer Verlag, 1990.

[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.

[10] S. Tominaga. Color image segmentation using three perceptual attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 628–630, 1986.

[11] 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.