Looking for Faces and Facial Features in Color Images

K. Sobottka I. Pitas

Department of Informatics, University of Thessaloniki 540 06, Greece

Phone: +30-31-996304 Fax: +30-31-996304

E-mail:{sobottka, pitas}@zeus.csd.auth.gr

Abstract

A first step towards face recognition is the localization of face-like regions and the extraction of facial features as eyes, mouth and nose. Although a lot of work has already been done in this research area, the recognition of human faces out of a scene with cluttered background is still a problem that deserves further investigation.

In this framework we present a novel approach to face localization by using color and shape information. For the extraction of facial features we employ two approaches. One is based on a modified watershed method and the second on min-max analysis. Results are shown for two different scenes.

Index Terms: Face recognition, color image processing, watersheds, min-max analysis.

1 Introduction

There are many different applications for face localization and recognition systems, e.g. model-based video coding, security systems, mug shot matching. Due to variations in illumination, background, visual angle and facial expressions, the problem is complex. A critical survey about existing literature on human and machine recognition of faces is given in [3].

The present paper deals with the topic of face localization and facial feature extraction. As it is described in [1], this approach is integrated in a multi-modal system for person verification using speech and image information. For the detection of facial regions and facial features, several approaches have been published so far using texture, depth, shape and color information or combinations of them. For example, in [4], the extraction of facial regions from complex background is done based on color and texture information. The input images are first enhanced by using color information, then textural features are derived by SGLD matrices and based on a textural model for faces, facial parts are detected. On the base of facial depth information primary facial features are located in [7]. In the first step pairs of stereo images containing frontal views are sampled from the input video sequence. Then point correspondences over a large disparity range are determined using a multiresolution hierarchical matching algorithm. Finally nose, eyes and mouth are located based on depth information. In [5] face localization is done by using shape information. An ellipse is chosen as model for the facial shape and based on edge information candidates for the head outline are determined. In order to extract facial features, in [11] first the input image is segmented using color characteristics, then feature points are detected and in the last step the different facial features are approximated by polynomials.

Most of the published approaches to face localization and facial feature extraction suffer either from their highly computational expenses or seem to be not very robust. Often edge information is used for facial feature extraction although facial features are not separated inside of the face by hard edges.

In this framework we present an approach for face localization using color and shape information. This combination of features allows a very robust face detection, because faces are significantly characterized by their skin color and oval shape. Facial feature extraction is done using greylevel information inside of the facial regions. Based on the observation that eyes and mouth differ from the rest of the face because of their lower brightness, we first enhance dark regions by applying morphological operations. Then facial features can be extracted on the base of watersheds or min-max analysis. We discuss both approaches and show results for two sample scenes.

2 Feature-based face recognition

In general the feature-based approach to 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 the first step of the face recognition process, facial regions are segmented out of the scene. Then facial features are extracted in the face regions. 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.

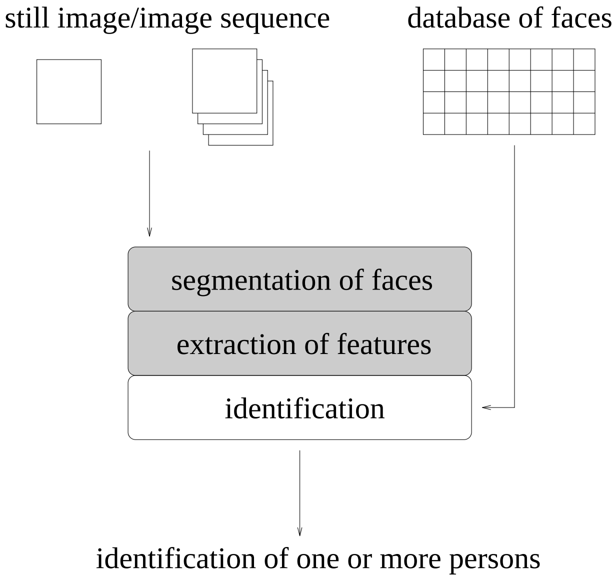


Figure 1: Problem of face recognition

Obviously, the order of the first two steps can be interchanged. For example in [2], 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 [8] an approach is presented which derives locations of whole faces from facial parts.

We have decided to segment firstly the face regions and then to extract facial features out of these facial regions. In case of color images, this processing order allows a very robust analysis, because faces differs significantly from the background by their color and shape. Thus this paper deals with the first two steps of the face recognition problem.

3 Localization of facial regions

Faces are significantly characterized by their oval shape and specific skin color. For that reason, we segment the facial regions on the base of shape and color information.

3.1 Segmentation based on color information

The effectiveness of using color information was already shown in [4] and [17]. Mostly the primary components of Red, Green and Blue are used for segmentation [15]. We have decided to consider the Hue-Saturation-Value (HSV) color space to extract skin color regions, because it is compatible to the human color perception. Alternatively the Hue-Saturation-Intensity (HSI) color space can be used as well, since it is very similar to HSV.

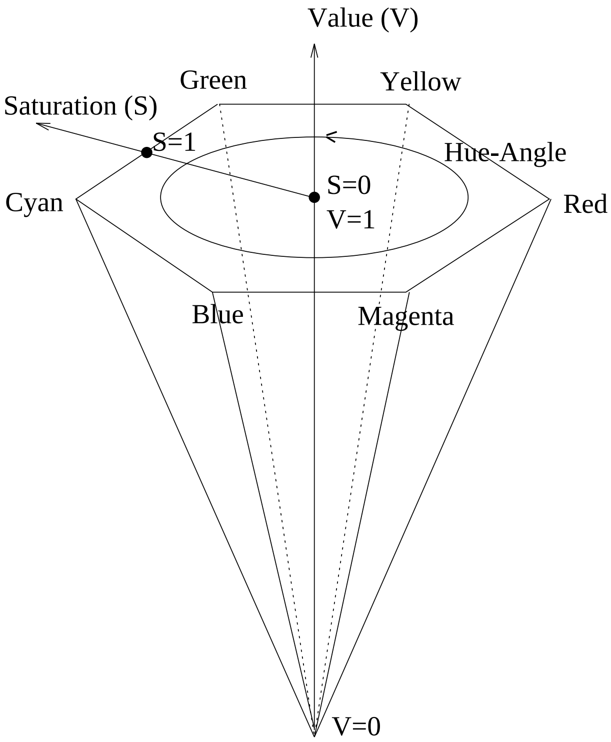


Figure 2: Hue-Saturation-Value color space

The HSV color space has hexcone shape as illustrated in Figure 2. Vertices of the hexagon at the top level represent Red, Yellow, Green, Cyan, Blue and Magenta. Neighbouring vertices are separated by a hue angle (H) of 60 degrees. Complementary colors are located opposite of each other. The saturation (S) defines the purity of colors and varies from 0 to 1. It increases with the radius of the hexagon inside of a level of the hexcone. The darkness of a color is specified by the value component (V), which varies from 0 and 1. The root of the hexcone is defined by $V = 0$ and the top level by $V = 1$. In case of $S = 0$, the value component describes the greylevel values with black at the root and white at the top level. A description of transformations from other color spaces into the HSV color space can be found in [13].

For the segmentation of skin-like regions it is sufficient to consider hue and saturation as discriminating color information. Besides by disregarding the value component, robustness towards changes in illumination and shadows is obtained. Skin segmentation in the Hue-Saturation space can be done by using appropriately defined domains of hue and saturation which describe the human skin color. As it is shown in Figure 3, this is equivalent to cut a sector out of the hexagon.

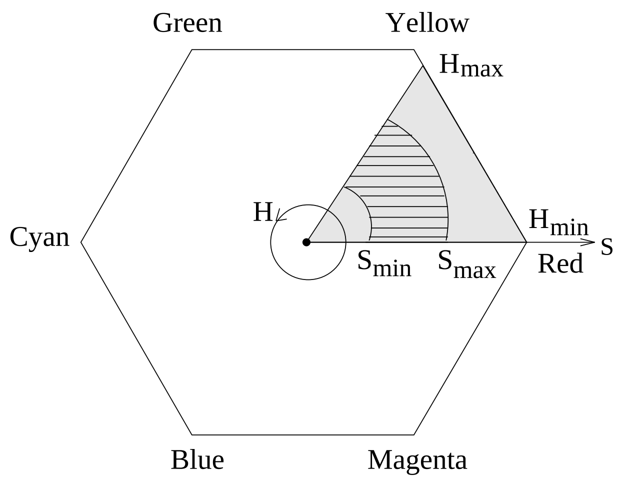


Figure 3: Thresholding in the Hue-Saturation space

The hue and saturation domains 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 you can see in Figure 4 these parameters are appropriate to segment the white skin as well as the yellow skin of human beings. The skin of hands and faces are well localized. There are some falsely detected pixels at the shirt, jacket and in the background.

(a) (b)

Figure 4: Results of segmentation based on color information: (a) color images and (b) segmented images

3.2 Connected component analysis

Because the face is more or less a connected region with skin color, we perform connected component analysis on the segmented image. We determine connected components by applying a region growing algorithm at a coarse resolution of the segmented image. Connectivity is determined on the base of the 4-pixel neighbourhood.

(a) (b)

Figure 5: Results of connected component analysis: (a) segmented image and (b) connected components

As you can see in Figure 5, we get rid of isolated falsely detected pixels in the background by this step. We obtain one large connected area for the face and another one for the hands. Several small connected components are found in the background. Based on shape information a further reduction of candidate regions is done.

3.3 Evaluation of shape information

The oval shape of a face can be approximated by an ellipse. Therefore looking for faces in images could be performed by detecting objects with nearly elliptical shape. This can be done based either on edges [5] or on regions, as we will show subsequently. The advantage of considering regions is that they are more robust against noise and changes in illumination. Thus we first compute for each connected component C the best-fit ellipse E on the base of moments [10]. Then we assess how well the connected component is approximated by its best-fit ellipse.

An ellipse is exactly defined by its center $(\overline{x}, \overline{y})$, its orientation θ and length a and b of its minor and major axis (Figure 6).

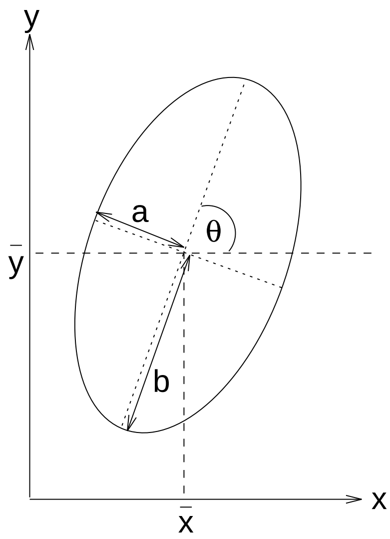


Figure 6: Parameter of an ellipse

The center $(\overline{x}, \overline{y})$ of the ellipse is given by the center of gravity of the connected component:

$$\overline{x} = \frac{1}{N} \sum_{(x,y)\in C} x \qquad \overline{y} = \frac{1}{N} \sum_{(x,y)\in C} y. \tag{1}$$

Here N denotes the number of pixels of the connected component C.

The orientation θ of the ellipse can be computed by using the central moments $\mu_{i,j}$ of the connected component:

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

The length of major and minor axis of the best-fit ellipse can be determined 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,

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

$$I_{max} = \sum_{(x,y)\in C} [(x - \overline{x})sin\theta - (y - \overline{y})cos\theta]^2, \tag{4}$$

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

$$a = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{max})^3}{I_{min}}\right]^{1/8} \tag{5}$$

$$b = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{min})^3}{I_{max}}\right]^{1/8}. \tag{6}$$

Then we assess, how well the connected component C is approximated by its best-fit ellipse E. For that purpose the following measure V is evaluated:

$$V = \frac{\sum_{(x,y)\in E}(1 - b(x,y)) + \sum_{(x,y)\in C\setminus E} b(x,y)}{\sum_{(x,y)\in E} 1} \tag{7}$$

with $b(x,y)$

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

denoting the indicator function of C. 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 which are outside of the ellipse (Figure 7).

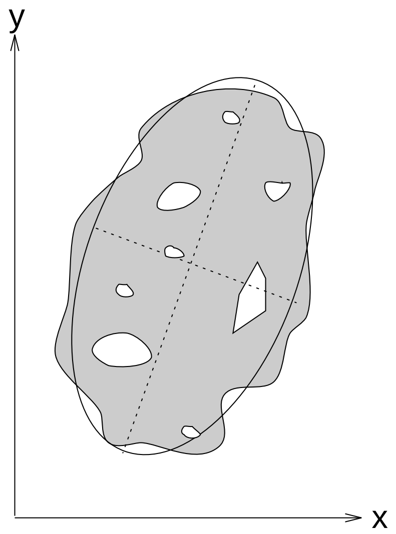


Figure 7: Approximation of a connected component by an ellipse

The ratio V 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, connected components that are well approximated by their best-fit ellipse are selected as face candidates (Figure 8).

On the base of the already computed elliptical parameters, a reduction of the number of 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. Subsequently the selected face candidates are verified by searching for facial features inside of the connected components.

Figure 8: Approximation of connected components by ellipses

4 Extraction of facial features

The approaches to extraction of facial features, presented in this paper, are based on the observation that, in intensity images, eyes and mouth differ from the rest of the face because of their lower brightness. This observation is also used in [14].

In case of eyes, the reason for that is the color of the pupils and the sunken eye-sockets. Even if the eyes are closed, the darkness of the eye sockets is sufficient to extract the regions of the eyes. The light red color of the lips emphasizes the mouth against its surrounding region and produces reduced local brightness in intensity images. This fact is more evident when the mouth is partially open.

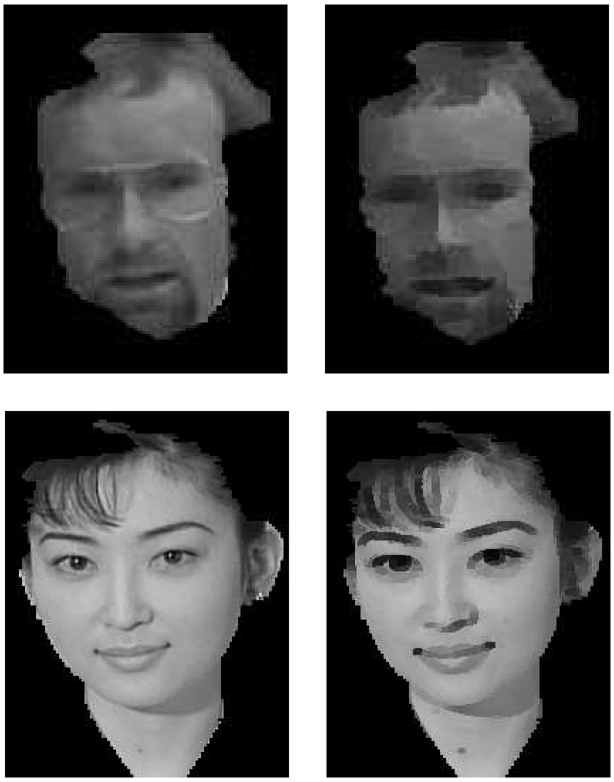
Therefore, in the following we consider the intensity information in the interior of the connected components (Figure 9a). In a preprocessing step, we enhance the dark regions inside of the connected components. For the extraction of facial features we discuss two approaches. The first uses a modified watershed method and determines the eye and mouth regions by flooding the greylevel relief. The second approach is based on min-max analysis and evaluates directly the x- and y-projections of the greylevel relief.

4.1 Enhancement of facial features

Enhancement of dark regions is done by using morphological operations. First we apply a greyscale erosion [9]. 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 interior of the connected components by the following extremum sharpening operation [12]:

$$g(x,y) = \begin{cases} min & \text{if } f(x,y) - min < max - f(x,y) \\ max & \text{otherwise} \end{cases} \tag{8}$$

Here min and max denotes the minimum and maximum greylevel 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 9b. Eyes and mouth and parts of hair and beard are emphasized.



(a) (b)

Figure 9: (a) Face regions and (b) enhanced facial features

4.2 Feature extraction using watersheds

Our approach to the extraction of eyes and mouth by watersheds is based on the following idea: The greylevel relief is immersed into water up to a predefined depth (Figure 10, black filled basins) and then the cutted minima are flooded by a modified watershed method (grey filled basins).

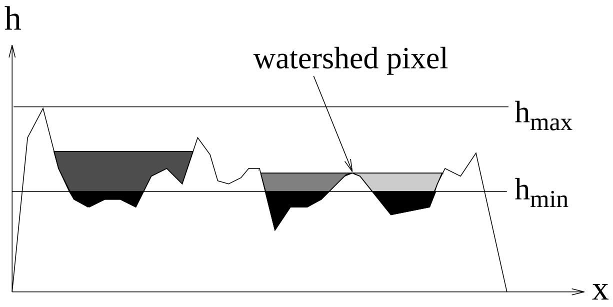


Figure 10: Modified watershed method

In case of our application this modified watershed method has the advantage over conventional watershed algorithms that only the minima of lower altitude are merged (Figure 10) and not a full segmentation is obtained. But the problem arises that, in contrast to conventional watershed algorithms, the merging process of a region is not stopped in any case by borders of other regions. This means that we have to control the merging process.

4.2.1 Immersion of greylevel relief

The first step, the immersion of the greylevel relief up to a predefined depth, is realized by thresholding. The threshold Θ_1 for a face candidate is determined in dependence on the minimum and maximum greylevel value in its interior. Then connected components are determined and different labels are assigned to pixels of different connected components. The result is considered as initial segmentation and we use it as input for our modified watershed method.

4.2.2 Modified watershed method

A very efficient realization of a conventional watershed algorithm is given in [16]. We use it partially to obtain also a very efficient modified version for our application. First the pixels of the interior of the face candidate region are sorted according to their greylevels. This ensures random access to the pixels and direct access to the neighbours. Beginning with the initial regions, determined by the thresholding step, the regions are slowly flooded. This is done by applying an iterative method. Let us denote the actual altitude by h and the altitude at the

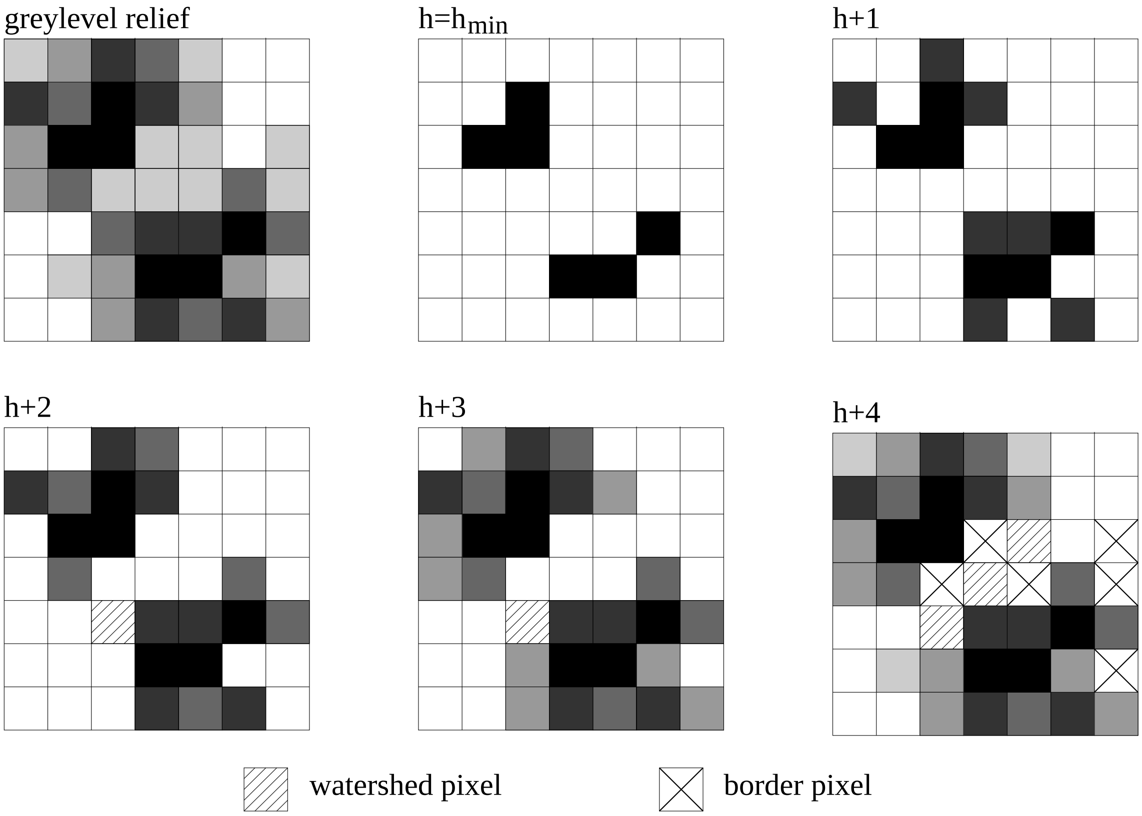


Figure 11: Flooding of cutted minima

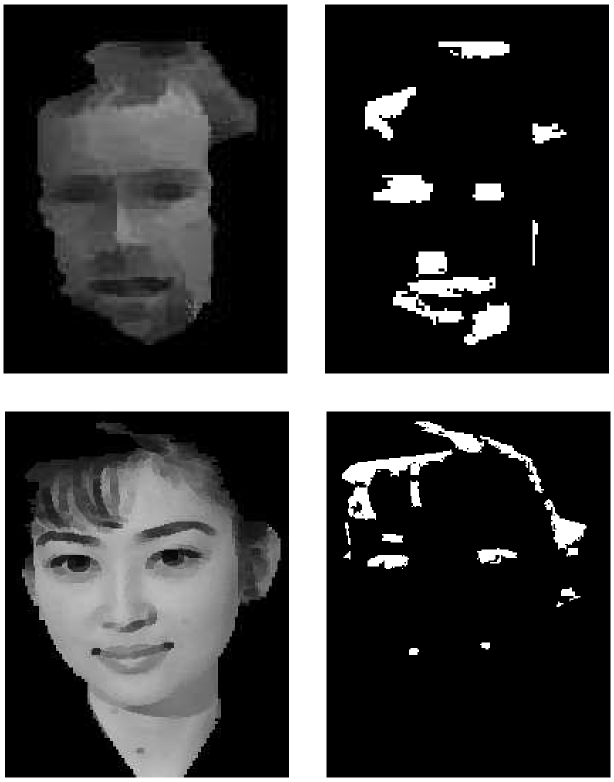
begin and end of the iterative process by h_{min} and h_{max} respectively. In our case, h_{min} and h_{max} are given by:

$$h_{min} = \Theta_1 + 1 \tag{9}$$

$$h_{max} = max_{(x,y)\in C}\ g(x,y), \tag{10}$$

with $g(x,y)$ denoting the preprocessed greylevel at position (x,y) and C denoting the connected component of the face candidate.

Starting with $h := h_{min}$, all pixels in the sorted array with greylevel h are directly accessed. For each of them we test, if it is in the neighbourhood of one of the already detected regions. If this is the case, the greylevel difference is computed between the two neighbours and, if the difference is under a predefined threshold, the pixel is merged to the already detected region. In the case that a pixel is in the neighbourhood of two or more regions, it is labelled as watershed pixel (Figure 10). If all pixels of an altitude h are processed, the next altitude $h := h + 1$ is considered and so on. The process stops when the maximal altitude is reached. An example for flooding is shown in Figure 11. Considering the local contrast ensures that the flooding process stops, also if there is no other region as direct neighbour. Such pixels are denoted as border pixels (Figure 11). Results for the two example scenes are shown in Figure 12.



(a) (b)

Figure 12: (a) Preprocessed face region and (b) extracted features using watersheds

In the first example, besides of regions for eyes and mouth, also regions for beard and hair are extracted. They could be removed by postprocessing. The second example shows a problem of the approach. The eyes are very well localized, but for the mouth only the corners are detected. The reason for that is that the lips are very bright in contrast to the corners of the mouth and the flooding process stops before merging the lips.

4.3 Feature extraction by min-max analysis

Facial features can also be extracted by another approach that is based on min-max analysis. In this case eyes and mouth are directly searched by evaluating the y-projection and x-projections of the topographic greylevel relief. Again we consider the preprocessed greylevel information as interior of the connected components (Figure 13b).

4.3.1 Normalization of orientation

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.

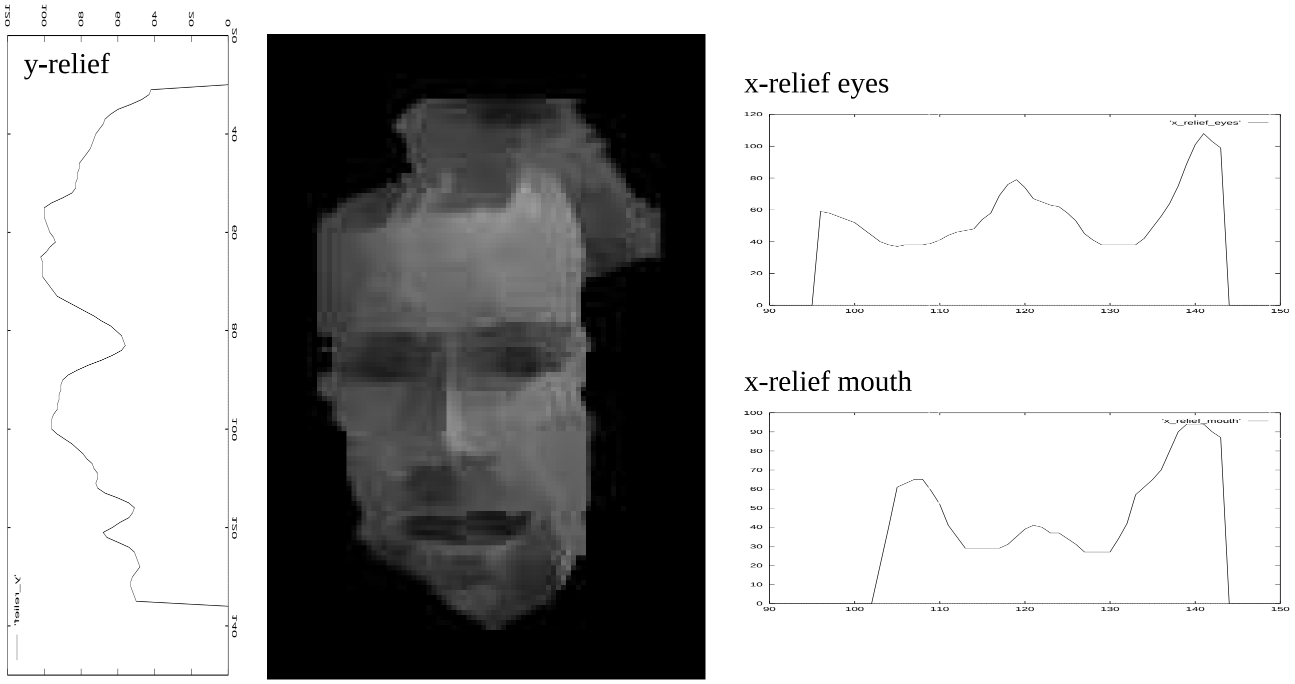
To obtain this normalization, a rotation of the interior of the connected component is necessary. The rotation angle is assumed to be equal to the orientation θ of the best-fit ellipse of the connected component. The coordinate transformation is defined by:

$$\begin{pmatrix} xr \\ yr \end{pmatrix} = \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{11}$$

Hereby xr and yr denote the rotated coordinates of x and y.

4.3.2 Computation of the y-relief and x-reliefs

After normalization of the orientation of a connected component the y-projection is determined. For that purpose we compute the mean greylevel of every row of the connected component. The resulting y-relief is smoothed by an average filter of width 3. Then minima and maxima are searched in the smoothed y-relief. By checking the gradients of each minimum to its neighboured maxima, significant minima are selected.



(a) (b) (c)

Figure 13: Min-max analysis

For each significant minima of the y-relief, x-reliefs are computed. To obtain more robustness, we determine x-reliefs for several rows around the minimum of the y-relief. Each x-relief is computed by averaging the greylevels of 3 neighboured rows of every column. Afterwards the resulting x-relief is smoothed in x-direction by an average filter of width 3. Then minima and maxima are determined.

As result, we obtain one smoothed y-relief (Figure 13a) for each facial candidate region with an attached list of its minima and maxima and, for each significant minima of the y-relief, smoothed x-reliefs (Figure 13c) with attached lists of their minima and maxima.

4.3.3 Search for eyes candidates

Beginning with the uppermost minima of the y-relief, we search through the lists of minima of x-reliefs to find two minima that meet the requirements for the positions of the eyes.

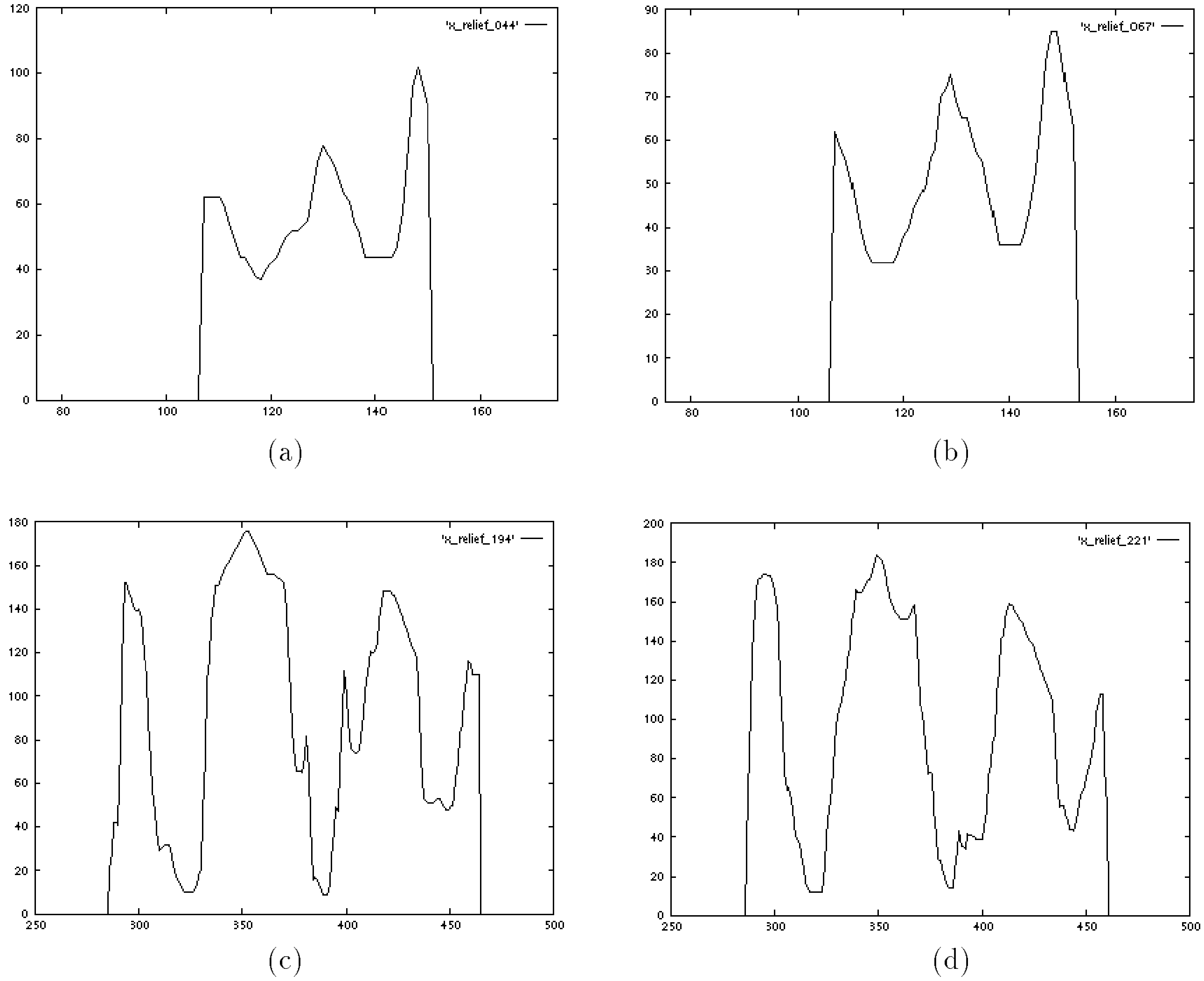


Figure 14: Examples of x-reliefs for the eye region

Typical examples for x-reliefs of the eyes region are shown in Figure 14. In Figure 14a,b examples of x-reliefs are given that contain only minima corresponding to the left and right eye. The examples in Figure 14c,d include also minima corresponding to the hair.

The requirements for eyes are specified as follows:

- The eyes are located in the upper or middle part of the head.
- There are significant minima for the left and the right eye.
- The minima have similar greylevel values.
- A significant maximum is between the two minima of the eyes.
- The ratio of eye distance to head width is within a certain range.

The assessment of how well a pair of minima meets the requirements, is done on the base of fuzzy theory. Thus we define a membership function for each of these requirements.

The membership function for the assessment of the ratio of eye distance to head width is shown as example in Figure 15a. 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. In the case that the measured ratio r is between $P2$ and $P3$, the requirement is entirely fullfilled and the certainty factor (CF) has the value 1.0. If $P1 < r < P2$ or $P3 < r < P4$ holds, the requirement is only partly fullfilled and $0.0 < CF < 1.0$. Otherwise the requirement is not fullfilled and $CF = 0.0$. Another example is the membership function for the assessment of similarity of greylevel values, illustrated in Figure 15b. With increasing greylevel difference of the left and right eye candidate, CF decreases. On the base of assessments for these criteria,

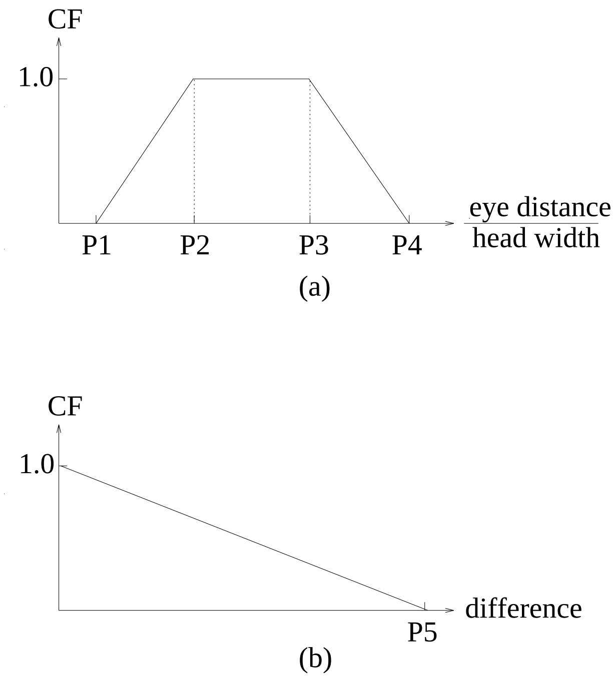


Figure 15: Assessment of criteria: (a) membership function for ratio of eye distance to head width, (b) membership function for similarity of greylevel values

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 eye candidates, represented by crosses, are shown in Figure 16.

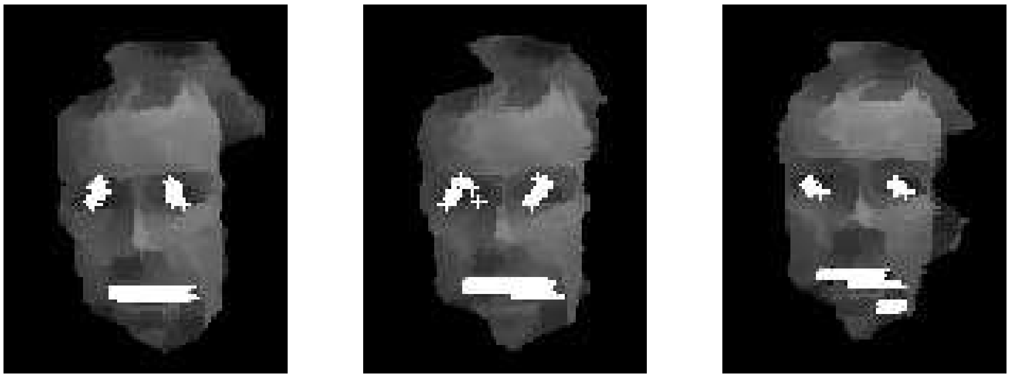


Figure 16: Eyes and mouth candidates

4.3.4 Search for mouth candidates

For the mouth a full search through all determined x-reliefs is done as well. As you can see in Figure 17, the mouth is characterized in x-reliefs by a significant basin with one or two minima inside. Mostly the corners of the mouth can be seen very well in x-reliefs (Figure 17c,d). Therefore we define the requirements for the mouth as follows:

- The mouth is located in the middle or lower part of the head.
- There are two significant maxima with no higher maximum between them.
- A significant minimum is between the maxima.
- The ratio of mouth width to head width is within a certain range.

In analogy to the search for eye candidates, every pair of maxima is assessed, on how well it meets the requirements for the mouth. As result we obtain for the face region a set of mouth candidates. Examples of mouth candidates, represented as horizontal line segments, are shown in Figure 16.

4.3.5 Selection of facial features

After the determination of eyes and mouth candidates, the best constellation of facial features is selected. For that we have to examine the cases shown in Table 1.

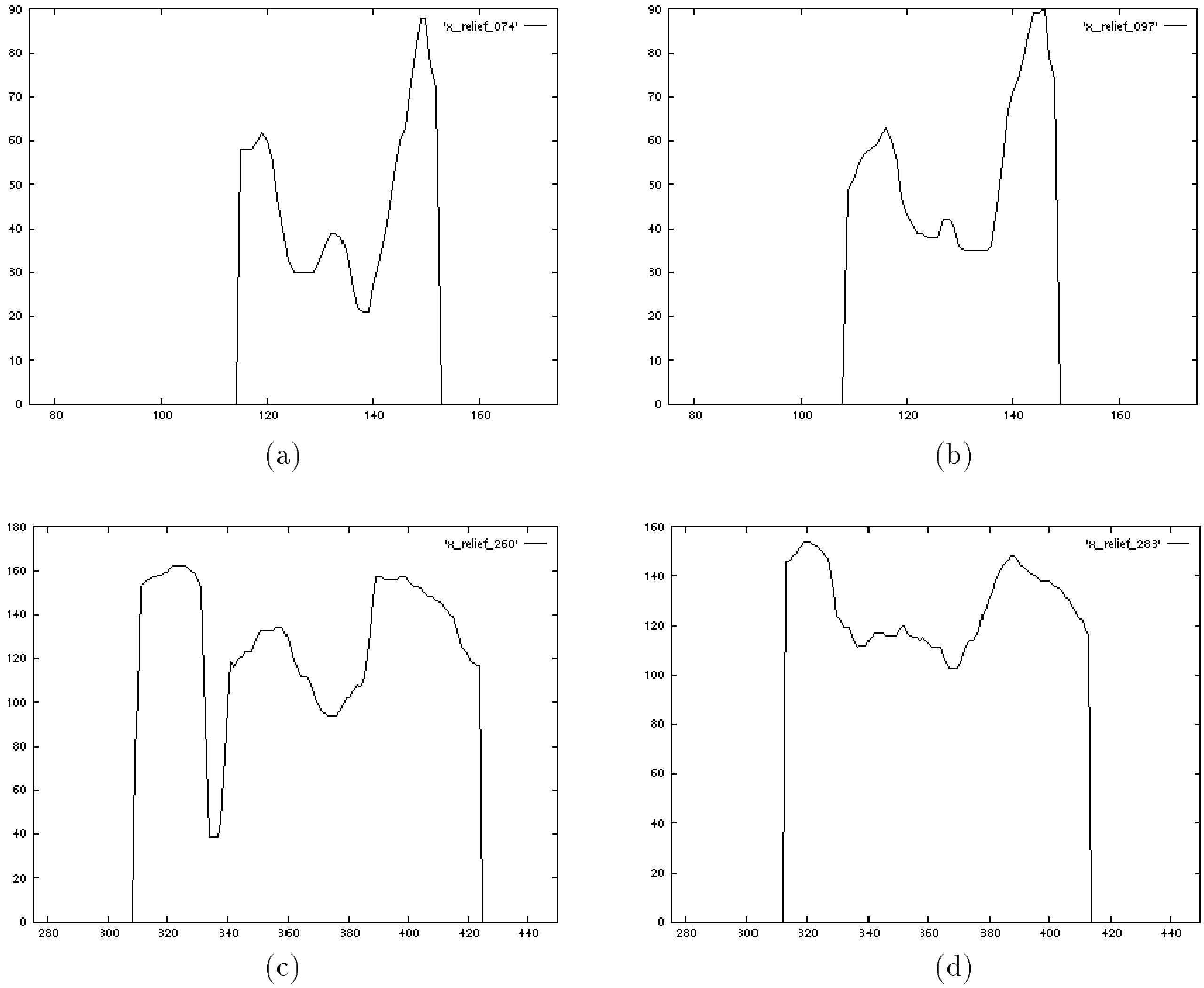


Figure 17: Examples of x-reliefs for the mouth region

case	eyes cand.	mouth cand.	selection
1	no	no	no facial features
2	yes	no	best cand. for eyes
3	no	yes	best cand. for mouth
4	yes	yes	best eyes-mouth constellation

Table 1: Cases for the selection of facial features

If we found no candidates for eyes and mouth (case 1), we have no facial features as result.

In the case, we detected only candidates for eyes (case 2) or only candidates for mouth (case 3), we chose the best candidate. For that we first cluster the candidates for eyes or mouth according to their x-coordinates of the left and right eye or left and right corner of the mouth. This is done using the unsupervised Min-Max-algorithm for clustering [6]. Starting with one pair of eyes or mouth corners as first cluster center, distances to all other candidates are computed. The candidate with maximum distance is chosen as second center. Then the following iterative procedure starts: For each candidate all distances to all cluster centers are computed and an attachment to the cluster with minimum distance is done. The candidate with maximum distance to its attached cluster is chosen as new cluster center. The iterative process stops, when this maximum distance is smaller than the half of the mean distance between cluster centers. As result we obtain a set of cluster centers. Each of them is a pair, consisting of the left and right center for eyes or mouth corners. Based on the number of votes for a center and the average assessment of candidates attached to the center, the best mouth or eyes candidate is chosen.

The best eyes-mouth constellation is determined, if eyes candidates as well as mouth candidates are detected (case 4). For that we first cluster the candidates of eyes and mouth as described in the previous paragraph. Then we build feature constellations by combining each cluster center of eyes with each cluster center of mouth. The best constellation is selected on the following criteria:

- The eyes candidate as well as the mouth candidate are well assessed.
- The ratio of vertical eye-mouth distance to head height is within a certain range.
- The mouth overlaps horizontally the region between left and right eye.

For the example scenes we obtain the results shown in Figure 18.

4.4 Comparison between approaches

Both presented approaches allow the extraction of facial features. The main advantage of feature extraction by the modified watershed method is that it is very efficient. No preprocessing is necessary and only the minima are flooded that are cutted by a thresholding step. The definition

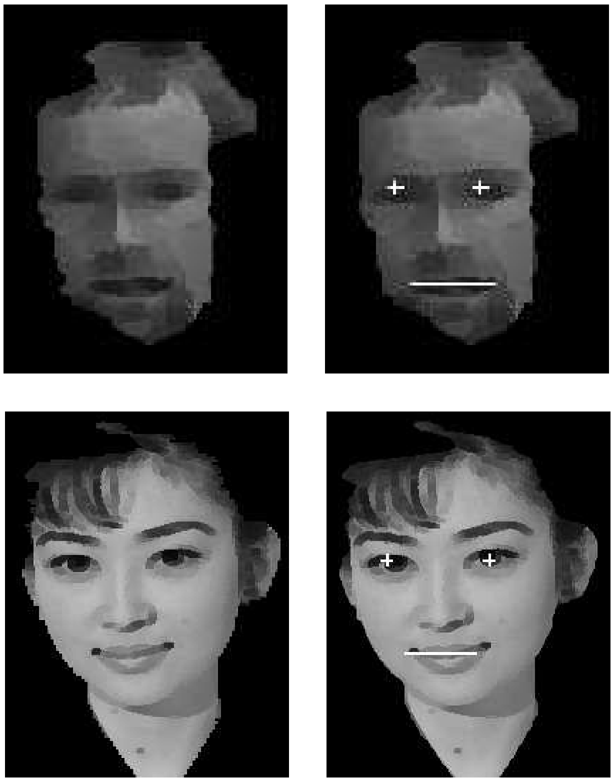


Figure 18: Results of facial feature extraction
by min-max analysis

of this threshold is critical, but can be chosen reliably in dependence on the minimum and maximum greylevel of the connected component. Problems in feature extraction arise because this method uses only local information during the flooding process. As shown in the second example in Figure 12 the flooding step stops before merging the lips of the woman.

By determining first the minima and maxima of the y-projection, the feature extraction based on min-max analysis employs a more global view. Only x-reliefs are evaluated, in which facial features are supposed. This effects also in a very efficient approach, though a normalization of orientation is necessary. Relationships between facial features can be assessed very well by using fuzzy theory. The advantages and disadvantages are summarized in Table 2.

4.5 Evaluation of facial feature extraction by min-max analysis

To assess the robustness of facial feature detection by min-max analysis, we applied our method to an image sequence consisting of 150 frames. Examples of frames are shown in Figure 19a,b

modified watersheds	min-max analysis
+ efficient − local − critical threshold	+ efficient + global − normalization of orientation

Table 2: Comparison of the approaches for facial features extraction

and Figure 20. 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 3. Facial features, that means eyes or mouth, are detected in 86% of

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

Table 3: Detection rates for facial features

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. For some frames faults occur. Examples of such faults are shown in Figure 19. For example in case of the mouth sometimes beard is mistaken as mouth (Figure 19a). Rarely the glasses are mistaken as eyes (Figure 19b). In the case of the image sequence shown in Figure 19c, sometimes problems with the eye brows arise, especially when the eyes are closed.

Further examples of results for both example scenes are illustrated in Figure 20 and Figure 21.

(a) (b)

(c)

Figure 19: Problems of facial feature extraction by min-max analysis

5 Conclusion

We have presented an approach which detects facial regions in color images on the base of color and shape information. For the extraction of facial features as eyes and mouth we have discussed two approaches.

Facial regions are determined using color and shape information. Therefore, we firstly extract regions with skin-like hue and saturation values and then compute the best-fit ellipse for each of the regions. Based on a function that assesses how well the regions are approximated by their

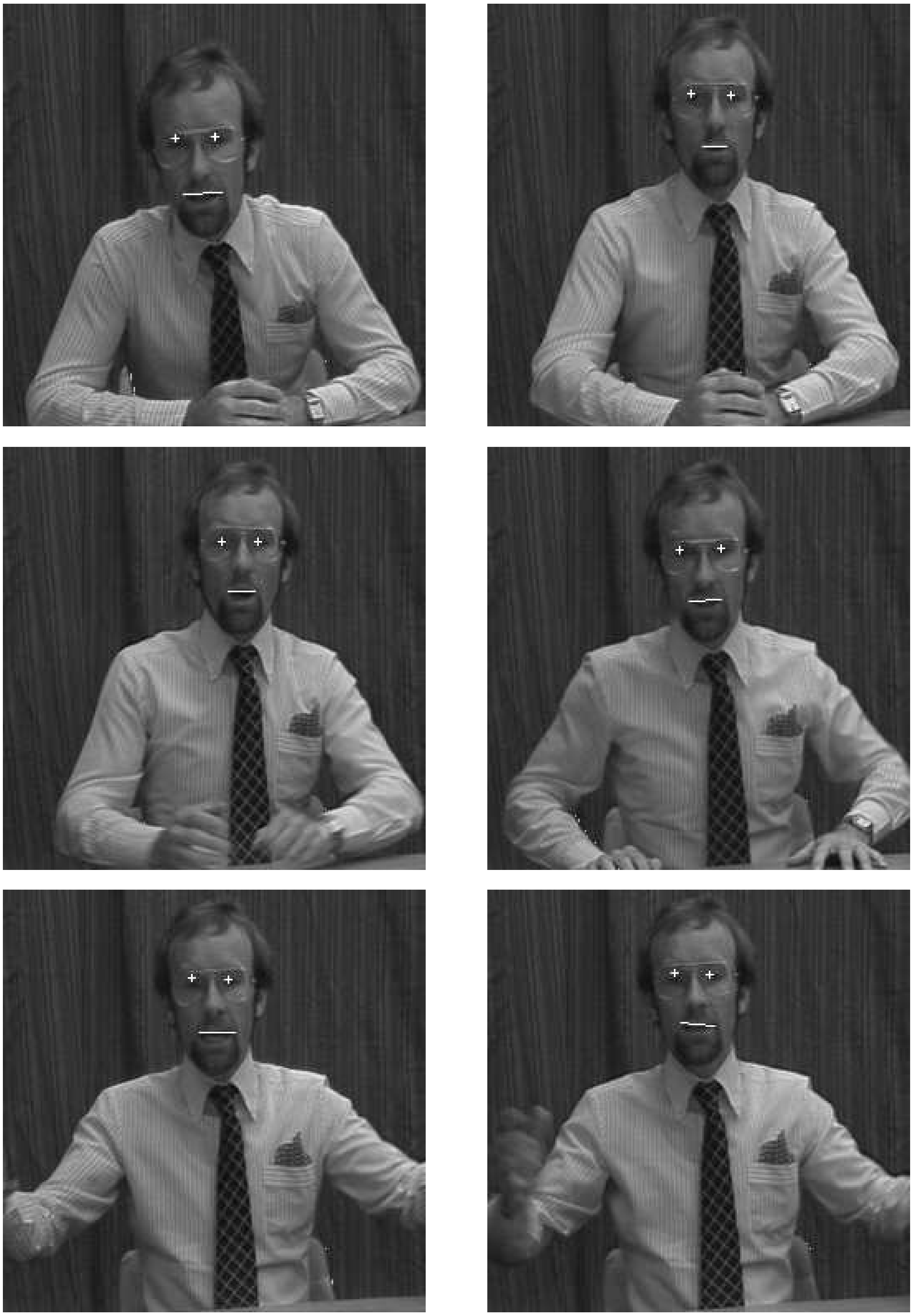


Figure 20: Results of facial feature extraction by min-max analysis

best-fit ellipse, we choose potential face candidates. These hypotheses for faces are verified by searching for facial features in the interior.

For the extraction of facial features two approaches have been presented. Both are based on the

Figure 21: Results of facial feature extraction by min-max analysis

observation that eyes and mouth differ from the rest of the face because of their lower brightness. In a preprocessing step, we enhance dark regions inside of the ellipses by applying morphological operations. In the first approach facial features are extracted by using thresholding and a modified watershed method. This works very efficiently, but in some cases problems arise because of the local view on the topographic greylevel relief. Feature extraction by min-max analysis employs more global characteristics. It is an efficient method, despite the fact that a normalization of orientation of face candidates is necessary. Results for both approaches were illustrated for different frames of two example scenes.

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, L. Vandendorpe. Multi-modal person verification tools using speech and images. *European Conference on Multimedia Applications, Services and Techniques*, Louvain-La-Neuve, Belgium, May 28-30, 1996.

[2] M. C. Burl, T. K. Leung and P. Perona. Face localization via shape statistics. In *International Workshop on Automatic Face and Gesture Recognition*, IEEE Computer Society, Swiss Informaticians Society, Swiss Computer Graphics Association et al., Zurich, Switzerland, June 26-28, 1995, pp. 154–159.

[3] R. Chelleppa, C. L. Wilson and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, vol. 83, no. 5, May 1995, pp. 705–740.

[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*, IEEE Computer Society, Swiss Informaticians Society, Swiss Computer Graphics Association et al., Zurich, Switzerland, June 26-28, 1995, pp. 238–242.

[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*, vol. 7, no. 3, July 1995, pp. 231–248.

[6] H. Ernst. *Einfuehrung in die digitale Bildverarbeitung.* Franzis, 1991.

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

[8] H. P. Graf, T. Chen, E. Petajan and E. Cosatto. Locating faces and facial parts. In *International Workshop on Automatic Face and Gesture Recognition*, IEEE Computer Society, Swiss Informaticians Society, Swiss Computer Graphics Association et al., Zurich, Switzerland, June 26-28, 1995, pages 41–46.

[9] A. N. Venetsanopoulos and I. Pitas. *Nonlinear Digital Filters: Principles and Applications.* Kluwer Academic Publishers, 1990.

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

[11] Y. Li and H. Kobatake. Extraction of facial sketch images and expression transformation based on faces. In *IEEE International Conference on Image Processing*, IEEE Computer Society Press, Washington D.C., USA, October 23-26, 1995, pages 520–523.

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

[13] I. Pitas. *Digital Image Processing Algorithms.* Prentice Halls, 1992.

[14] M. J. T. Reinders, P. J. L. van Beek, B. Sankur and J. C. A. van der Lubbe. Facial feature localization and adaption of a generic face model for model-based coding. *Signal Processing: Image Communication*, vol. 7, no. 1, Jul 1995, pp. 57–74.

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

[16] L. Vincent and P. Soille. Watersheds in digital space: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, June 1991, pp. 583–598.

[17] H. Wu, Q. Chen and M. Yachida. An application of fuzzy theory: Face detection. In *International Workshop on Automatic Face and Gesture Recognition*, IEEE Computer

Society, Swiss Informaticians Society, Swiss Computer Graphics Association et al., Zurich, Switzerland, June 26-28, 1995, pp. 314–319.