A Fully Automatic Approach to Facial Feature Detection and Tracking

Karin Sobottka and Ioannis Pitas *

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

Abstract. The detection of facial features is a necessary step for a wide range of applications e.g. person verification, lipreading and model-based face coding. Due to changes in illumination, visual angle and facial expressions, the variability of facial features in their appearance is high. A robust approach to facial feature detection has to handle these variations. In this framework, we present an approach for the extraction of eyebrows, eyes, nostrils, mouth and chin. Our approach for facial feature extraction is based on the observation that facial features differ from the rest of the face because of their low brightness. Thus, we detect facial feature candidates by evaluating the topographic greylevel relief of the face region. Based on vertical symmetry, distances between facial features and the assessment of each facial feature, we choose the best face constellation. Incomplete face constellations are considered as well. Once facial features are detected in an image sequence, they can be tracked over time. We perform facial feature tracking by block matching. The best-match position is refined by minima analysis. The success of our approach was tested on 38 different image sequences containing faces.

1 Introduction

Several approaches for facial feature extraction are proposed in the literature. Some of them derive the locations of facial features directly out of the front-view images (e.g. [2], [7],[6]). But more often, first a segmentation of facial regions is done and then facial features are searched inside of these regions (e.g. [4], [5]). An excellent survey on human and machine recognition of faces is given in [3]. Also [1] gives a very good overview about existing approaches and techniques in this research field.

The present paper proposes a novel approach for facial feature extraction. The segmentation of facial regions is done in advance by using color and shape information ([8], [9]). We employ eyebrows, eyes, nostrils, mouth and chin as interesting facial features. Facial features are localized by evaluating the topographic greylevel relief of facial regions. Once facial features are reliably detected, we track them over time. Tracking is performed by searching for corresponding facial feature blocks in consecutive frames. In order to ensure the right block content, the greylevel relief is evaluated for each facial feature block.

2 Detection of Facial Features

Our approach for facial feature detection is based on the observation that facial features differ from the rest of the face because of their low brightness. For

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

example, in the case of the eyes, this is due to the color of the pupils and the sunken eye-sockets. Even if the eyes are closed, the darkness of the eye sockets is sufficient for eye extraction. Therefore, we shall employ the intensity information in the interior of facial regions.

2.1 Face Model

In our approach, a face can be described by a constellation of minima and maxima of the topographic greylevel relief. Because of the similarity of some of the facial features in their appearance, we subdivide facial features in three groups. The first group contains eyes and eyebrows, the second group describes the nostrils and the third group characterizes mouth and chin. For each group of facial features a description is defined based on minima and maxima and relative head position (Table 1).

group 1: eyebrows, eyes	group 2: nostrils	group 3: mouth, chin
two significant minima	two significant minima	two significant maxima
upper/middle part of head	middle part of head	middle/lower part of head
significant maximum between minima	significant maximum between minima	significant minimum between maxima
ratio of distance between minima to head width is in certain range	small distance between minima	ratio of distance between maxima to head width is in certain range
similar greylevels		

Table 1. Description of facial feature groups

Based on this rough description of facial feature groups, a face constellation can be defined as follows:

$$\text{Face constellation} = [cand_1]^*[cand_2][cand_3]^* \tag{1}$$

where $cand_i$ denotes candidates for group i. Equation (1) defines that a face constellation may consist of a sequence of candidates for group 1, one candidate for group 2 and a sequence of candidates for group 3 (the star in (1) denotes a sequence of candidates). By considering each part of the face constellation as optional, the face model takes into consideration possible incomplete face constellations. The maximal number of candidates for group 1 and group 3 belonging to a face constellation is limited to two, due to the fixed number of facial features inside of a human face.

2.2 Extraction of Feature Candidates

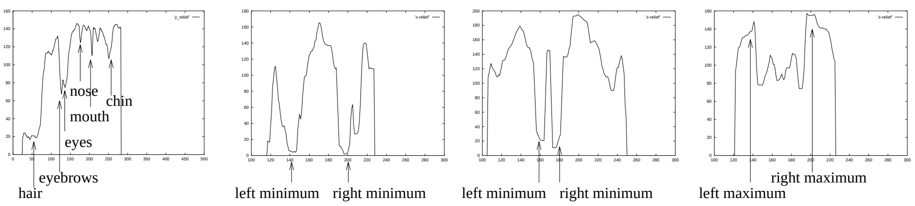
In a preprocessing step, we enhance dark regions in the interior of facial regions by using morphological operations [9].

Then candidates for each of the groups are determined by searching for minima and maxima in the topographic greylevel relief. This can be done by using watersheds [10], or as we shall show here, by directly evaluating the x- and y-projections of the greylevel relief.

We compute the y-projection of the topographic greylevel relief by determining the mean greylevel of every row of the face region. Then we smooth the y-relief by an average filter of width 3 in order to get rid of small variations in the y-relief. A typical example for an y-relief is illustrated in Fig. 1a. Significant minima can be seen for hair, eyebrows, eyes, nose, mouth and chin.

Significant minima are determined in the y-relief by checking the gradients of each minimum to its neighbour maxima. Each significant minima is considered as possible vertical position for facial features. Therefore, more detailed information is computed along the horizontal direction. Thus, beginning with the uppermost minima in y-direction (leftmost minima in Fig. 1a), we determine x-reliefs by averaging the greylevel values of 3 neighbour rows of every column. Afterwards the resulting x-reliefs are smoothed in x-direction by an average filter of width 3 and minima and maxima are determined. As a result, we obtain for a face region one smoothed y-relief with an attached list of its minima and maxima and, for each significant minima of the y-relief, smoothed x-reliefs with attached lists of their minima and maxima.

By searching through the lists of minima and maxima, candidates of the three facial feature groups are determined. A candidate for group 1 is found, if two significant minima in x-direction are detected that meet the requirements of group 1 (see Table 1). An example of a x-relief containing a candidate for group 1 is shown in Fig. 1b.



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

Fig. 1. Examples for (a) y-relief and x-reliefs containing a (b) group 1, (c) group 2, (d) group 3 candidate.

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 the requirements [9]. On the basis of these assessments, candidates for group 1 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.

After candidates are determined, we cluster them according to their left and right minimum x-coordinates. This is done using the unsupervised Min-Max-algorithm for clustering [10]. By this step, we reduce the number of candidates significantly and obtain representative candidates for group 1.

In analogy, the search for candidates for group 2 and group 3 is done. Examples of x-reliefs containing candidates for group 2 and group 3 are illustrated in Fig. 1c,d.

2.3 Best Face Constellation

After candidates for the facial feature groups are detected, the best constellation of facial features is determined. For that we build all possible face constellations and assess each of them based on the vertical symmetry of the constellation, the distances between facial features and the assessment of each facial feature candidate. Incomplete face constellations are taken into consideration by allowing void facial feature positions. In order to enforce the preference to complete rather than less complete face constellations, the assessment of a face constellation is multiplied by a weighting factor, which depends on the number of facial features belonging to the face constellation. According to the assessment of the constellations, the constellations are ranked and the best constellation is chosen.

2.4 Results on Facial Feature Detection

To assess the robustness of our approach, we applied our method to the European ACTS project M2VTS database. The database includes 37 different faces containing features as beard, glasses and different facial expressions. Results of the detection of facial features are shown in Fig. 2. Eyebrows are represented by two short horizontal line segment, eyes by crosses, nostrils by small squares, the mouth by a horizontal line segment and the chin by a elongated rectangle. The results of evaluating the facial feature detection are illustrated in Table 2.

features	detected [%]	correctly [%]	falsely [%]
eyebrows	81	83	17
eyes	94	74	26
nostrils	94	97	3
mouth	92	97	3
chin	86	97	3

Table 2. Detection rates for facial features extraction.

These results are rather satisfactory. Detection errors for eyebrows and eyes tend to correlate with each other. Sometimes eyebrows are detected at the position of the eyes. This problem arises particularly in cases, when only eyes or eyebrows are extracted. In those cases the contextual information does not help distinguishing eyebrows from eyes. Problems in chin detection arise, if a person has a full beard. Such cases need special attention for facial feature detection.

3 Tracking of Facial Features

If facial features are reliably detected once, they can be tracked over time in an image sequence. Up to now facial feature tracking is only realized for eyes and mouth. But in analogy to the proposed method also tracking of eyebrows, nostrils and chin can be performed. Relatively robust tracking is possible by using block matching. First, initial blocks have to be extracted. Then they can be tracked over time by searching for corresponding blocks in consecutive frames. To ensure that the reference blocks always contain facial features, we verify the block contents by minima analysis.

3.1 Initialization of Feature Blocks

The initial block size is defined in the case of the eyes in dependence on the eye distance and, in case of the mouth, in dependence on the mouth width. Let us

Fig. 2. Results of face segmentation and facial feature detection.

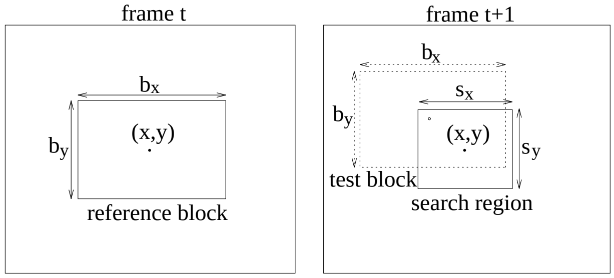
denote the block width and block height of a feature block B by b_x and b_y. Then the initial block for a feature is defined by placing the center of the block frame with dimension b_x and b_y on the detected feature position (x, y) at time t.

3.2 Matching of Feature Blocks

When initial feature blocks at time t are extracted, the position of features at time $t+1$ can be determined by matching the blocks of time t to the frame of time $t+1$ (Fig. 3). Because the displacement between consecutive frames is restricted due to video rate, the search for corresponding blocks can be restricted to a search region of dimension s_x and s_y centered at the feature position (x, y) at time $t+1$ (Fig. 3b).

For each position of the search region a similarity measure between reference block and test block is evaluated. In our case we use the sum of absolute greylevel differences as block difference BD. For a displacement (u, v) between reference block and test block, the block difference at position (x, y) results in:

$$BD_{x,y}(u, v) = \sum_{i,j \in B_{x,y}} |f(i, j, t) - f(i+u, j+v, t+1)| \tag{2}$$



(a) (b)

Fig. 3. Block matching (a) frame t with reference block and (b) frame $t+1$ with search region and one of the test blocks.

where $f(i, j, t)$ denotes the greylevel value at position (i, j) at time instance t. The best-match position is found, if the block difference is minimal.

Matching results are illustrated in Fig. 5. The search regions for corresponding blocks are shown for eyes and mouth. The greylevel values inside of the search regions represent the quality of block matching: the higher the greylevel value the better the matching.

3.3 Refinement and Updating

After performing block matching we refine the best-match position by minima analysis. In analogy to the extraction of facial features, we determine minima in the y-relief of the best-match block and for each minima, we determine minima in the related x-reliefs. After minima localization we select the minimum that is closest to the best-match position and consider its position as refined feature position. By this refinement step we ensure that the best-match block at time t still contains a facial feature as content.

Facial feature block updating is critical for tracking. Once a fault occurs, the whole tracking process collapses and a new face recognition is necessary. Thus, it has to be ensured that block updating is done correctly. In our approach, we ensure this by evaluating the best-match position and the refined feature position. If the distance between best-match position and refined position is small, we define a new reference block at the refined feature position. Otherwise, we consider the match as uncertain and the reference block remains constant.

3.4 Results on Facial Feature Tracking

Results of facial feature tracking for an image sequence consisting of 150 frames are shown in Fig. 4 and 5 for every tenth frame. The reference blocks extracted for the right eye and mouth for every tenth frame are illustrated in Fig. 4. Though eye and mouth close and open during tracking, reference blocks contain always the right content.

The search regions for facial features and the results of block matching for every tenth frame are shown in Fig. 5.

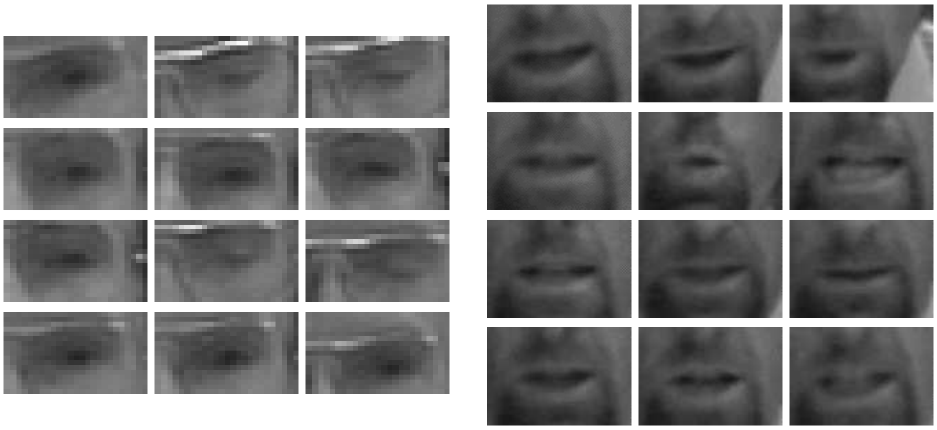


Fig. 4. Reference blocks for every tenth frame of right eye and mouth.

4 Conclusion

In this paper we have presented a fully automatic approach for facial feature extraction and tracking. Results were shown for several images containing faces.

References

1. 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.
2. 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.
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. 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.
5. 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.
6. 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.
7. B. S. Manjunath, R. Chellappa, C. Shekhar, and C. von der Malsburg. A robust method for detecting image features with application to face recognition and motion correspondence. In *11th IAPR International Conference on Pattern Recognition*, pages 208–212, The Hague, The Netherlands, August 30 –September 3 1992. IEEE Computer Society Press.
8. K. Sobottka and I. Pitas. Extraction of facial regions and features using color and shape information. In *Int. Conf. on Pattern Recognition (ICIP)*, Vienna, Austria, August 1996.
9. K. Sobottka and I. Pitas. Face localization and facial feature extraction based on shape and color information. In *Int. Conf. on Image Processing (ICPR)*, Lausanne, Switzerland, September 1996.

Fig. 5. Results of facial feature tracking: Search regions for corresponding feature blocks and quality of block matching represented as greylevel intensity inside the search regions (the smaller the block difference the higher the greylevel value).

10. K. Sobottka and I. Pitas. Looking for faces and facial features in color images. *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, Russian Academy of Sciences*, 1996.