Edge and Keypoint Detection in Facial Regions*

R. Herpers[1], M. Michaelis[1], K.-H. Lichtenauer[1], and G. Sommer[2]

[1]GSF – Institut für Medizinische Informatik und Systemforschung, MEDIS
Ingolstädter Landstr. 1, D–85764 Oberschleißheim, Germany
Email: herpers@gsf.de
[2]Institut für Informatik, Lehrstuhl für Kognitive Systeme
Christian-Albrechts-Universität, Preußerstr. 1-9, D–24105 Kiel, Germany

Abstract

In this contribution we introduce a method for the automatic detection of facial features and characteristic anatomical keypoints. In the application we are aiming at the anatomical landmarks are used to accurately measure facial features. Our approach is essentially based on a selective search and sequential tracking of characteristic edge and line structures of the facial object to be searched. It integrates model knowledge to guarantee a consistent interpretation of the abundance of local features. The search and the tracking is controlled in each step by interpreting the already derived edge and line information in the context of the whole considered region. For our application, the edge and line detection has to be very precise and flexible. Therefore, we apply a powerful filtering scheme based on steerable filters.

Keywords: Keypoint and feature detection, knowledge integration, face recognition, steerable filters

1. Introduction

Facial keypoints such as eye corners and mouth ends are important features for many different tasks in automatic face processing [4, 5]. The localization of facial keypoints is usually performed interactively or it is not very precise and robust. In general, the problem is that anatomical facial landmarks we are searching for in this paper are defined rather as morphological features (e.g. the corner of an eye) than by a low-level definition only based on the image data. Different realizations of the same facial keypoint can vary drastically in terms of their grey value distribution in the image. Hence, it is not possible to detect facial keypoints by standard corner detectors or other purely local and data-driven detectors that do not make use of the context or of appropriate model knowledge. The problem of local and data-driven keypoint detectors is that there is too much local image structure in complex real world images. Therefore, it is not possible to give any interpretation to local features without considering their context.

Our approach starts by detecting the most prominent and reliable features (edges and lines) in the considered facial region. For a robust detection and a subsequent tracking of the edges and lines we propose a filtering scheme that is based on steerable filters [1, 6, 7]. The line and edge detectors can be calculated very efficiently in any orientation and scale. This flexibility is very important for the sequential search strategy we have developed. For the detection of the facial keypoints we have defined three 'basic filter operations', of which the sequential search strategy is composed and which are reused over and over again for every distinct facial region and for every distinct keypoint.

The application, that gives us the motivation, is the detection of dysmorphic facial signs. Dysmorphic signs in face images are minor anomalies which, by definition, do not lead to functional disturbances [9] (fig. 1).

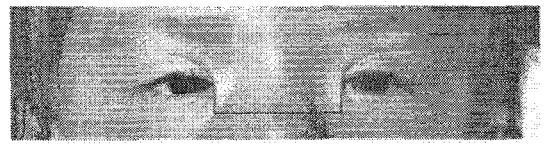


Figure 1. Example of a very enlarged intercanthal distance (taken from [10, p. 42]).

In this context, dysmorphic signs in faces play an important role in syndrome identification because in most cases they are visible and prominent in the facial image [8]. It has been shown, that particular phenotypic combinations are typical for distinct dysmorphic syndromes [9, 10]. Therefore, the detection of particular keypoint positions (fig. 2) in dysmorphic face images is of a high diagnostic value. For this, the localization of the several keypoints should be very accurate, reproducible, and it should correspond to the anatomical definition of the keypoint positions.

*This work is partially supported by the DFG grants So 320\1-1 and Ei 322\1-1

0-8186-7713-9/96 $5.00 © 1996 IEEE

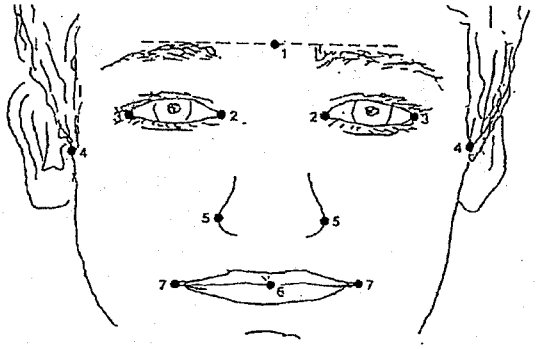


Figure 2. Important keypoints in a frontal face image. The figure is taken from [8, p. 63].

2. The filtering scheme

The filtering scheme is based on line and edge detectors. For this, we use a first and second derivative of Gaussian:

$$F_1(x,y) = C_1 \, \frac{y}{\sigma} e^{-\frac{x^2}{2(\varepsilon\sigma)^2}} \, e^{-\frac{y^2}{2\sigma^2}} \qquad (1)$$

$$F_2(x,y) = C_2 \, \left(1 - \left(\frac{y}{\sigma}\right)^2\right) e^{-\frac{x^2}{2(\varepsilon\sigma)^2}} \, e^{-\frac{y^2}{2\sigma^2}} \qquad (2)$$

C_1 and C_2 are normalization constants such that both filters have the same L^2 norm (for F_1 see fig. 3 left). These filters are steered in orientation and scale (ϑ,σ) to be applied separately as line and edge detectors. The term 'steerability' refers here to a linear reconstruction of the deformed filters $F_{\vartheta,\sigma}$ on the base of a small number N of so called basis functions $A_k, k = 1 \ldots N$ (see superposition formula 3). Typically N will be small (10 or 20), while ϑ and σ theoretically assume an infinite number of values. For more details see [6] and [1, 7].

$$F_{\vartheta,\sigma}(\vec{x}) \approx \sum_{k=1}^{N} b_k(\vartheta,\sigma) A_k(\vec{x}) \qquad (3)$$

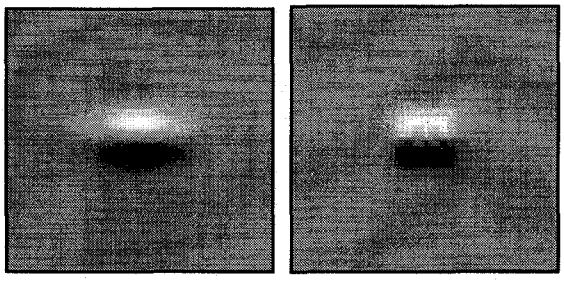


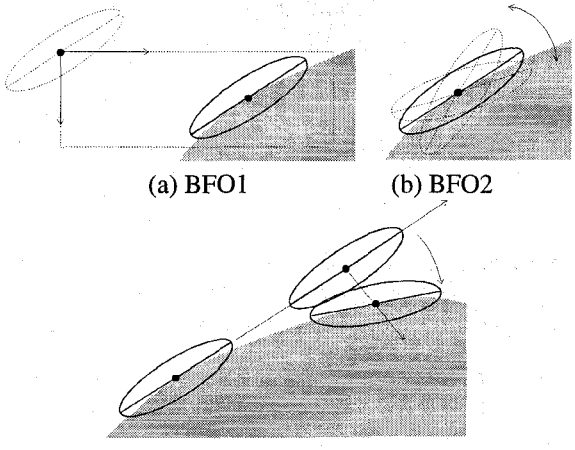
Figure 3. Edge detection filter F_1 (first derivative of Gaussian, with an aspect ratio of 2) in original resolution (left) and reconstructed (right) with 10 basis functions.

The quality of the reconstructed primitive filters depends on the number of used basis functions. Figure 3 shows the original (left) and the reconstructed (right) edge detection filter F_1 using only 10 basis functions. The following two properties of the basis functions are essential for the design of our filtering scheme because they allow for an easy on-line adaptation of the tradeoff between the speed and the quality of the filters.

- The basis functions are orthogonal. Thus, it is easy to add on-line new basis functions to achieve a better reconstruction quality.
- Any number of basis functions reconstruct **all** deformed filters. Only the quality of the reconstruction changes.

In most cases low quality approximations of the used filters are sufficient because they qualitatively still resemble elongated edge and line detectors. The detection and tracking of the keypoints is performed by three different **basic filter operations** that make extensively use of the deformed filters in arbitrary orientations and scales.

- The first operation (BFO1) searches in a predefined region for an edge or line with a predefined orientation and scale (fig. 4a).
- The second operation (BFO2) determines the orientation of an edge or line at a given position by evaluating the maximal response of a rotated filter (fig. 4b).
- The third operation (BFO3) tracks an edge or line by one step. For this, the filter is moved by a small step in the already known direction. Then the edge or line is searched again in perpendicular direction (fig. 4c).



(a) BFO1 (b) BFO2

(c) BFO3

Figure 4. Basis filter operations. BFO1: Detection of an edge or line (a). BFO2: Determination of the orientation (b). BFO3: Stepwise tracking of an edge or line (c).

3. Context based keypoint detection

We have already pointed out that the integration of model knowledge is mandatory. The applicability of any model of course is restricted to a certain class of images. Therefore, the facial images studied in this contribution have to fulfill several general requirements to be accepted for the processing. The general requirements comprise the necessary

resolution (512^2 full face image), the orientation (frontal, not tilted faces), the illumination (frontal and diffuse), and the completeness (no occlusions or wearing glasses) of the facial images. Furthermore, there are some detailed assumptions. The detailed assumptions are essentially used to control the sequential search strategy. They consist of particular model knowledge about the image structure of the considered region. They describe the kind, the existence, and the orientation of different prominent structures and keypoints in the region such as a prominent, curved, bright-to-dark edge segment of the iris with a nearly vertical orientation. A detailed description of the complete model is beyond the scope of this paper, but more details can be found in [3]. The model of the edge structures of a left eye which are essentially considered, detected, or tracked during the processing are depicted in figure 5.

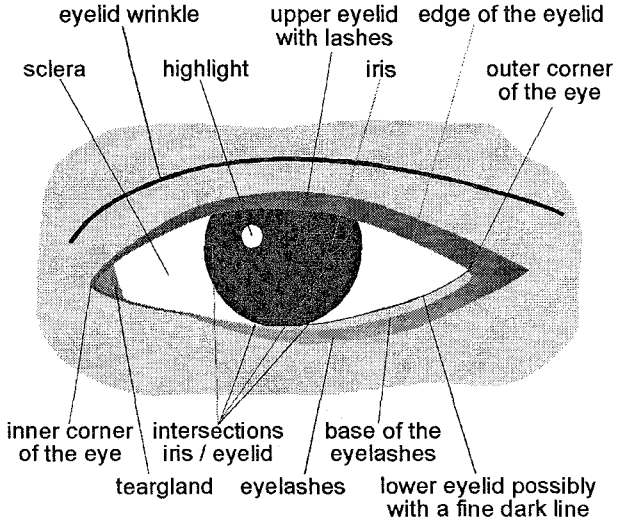


Figure 5. Model of a left eye. The depicted details in the model are important features for the sequential detection and tracking.

We want to emphasize that the presented approach and the class of images are given by our medical application which is different from general face processing situations (e.g. resolution etc.). However, it is straight forward to adapt the developed method to other situations and to any class of facial images.

4. Sequential search strategy

The detection of the different keypoints involves several steps. Each step implies several applications of the **basic filter operations** (fig. 4). The selection of the operations and their parameters for each step is determined and controlled by the already derived information together with the model. From these combined filter operations a sequential search strategy is developed to detect the different characteristic features.

To better understand the principle of the sequential search strategy, the detection of the iris and of the eye corners will be described in detail (fig. 6).

The most prominent and reliable features within the eye region are the edges of the iris (fig. 6 upper left). Therefore, the sequential search starts by detecting the left edge segment (vertical, bright-to-dark step edge) of the iris applying the basis filter operation BFO1 (fig. 4a). The detected edge is tracked upwards and downwards (using BFO3) until the intersection points with the eyelids are reached. Subsequently the corresponding right edge segment of the iris is searched using the model knowledge and the information about the size, position and orientation of the already detected left edge segment. Finally the iris is segmented and the center and the radius of the iris is determined (fig. 6 upper right). The computed features are checked with the model knowledge concerning to the expected circular symmetry, size of the radius, and some other constraints.

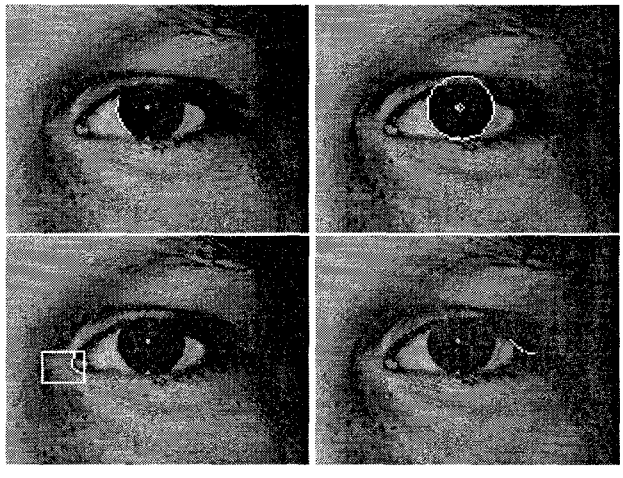


Figure 6. Example steps of the sequential search for a left eye. First, a prominent vertical bright-to-dark edge is detected. After the detection of the corresponding right edge segment, the final segmentation of the iris is computed (first row). The eyelid edges are searched, tracked, and finally that edge segment which is strongly curved is detected to determine the inner eye corner. For the outer eye corner, the upper and lower eyelid edges are tracked until they end (second row).

After the detection of the intersection points of the iris and the eyelids, the edges of the eyelids are tracked until an edge segment which is strongly curved indicate the inner eye corner (fig. 6 lower left) respectively the intersection of the tracked edges marks the outer eye corner (fig. 6 lower right). The tracking is controlled at each step by the integrated model knowledge (fig. 5) and by the actually derived information to avoid e.g. that the tracking is misled by edges of the tear gland. The procedure presented has the

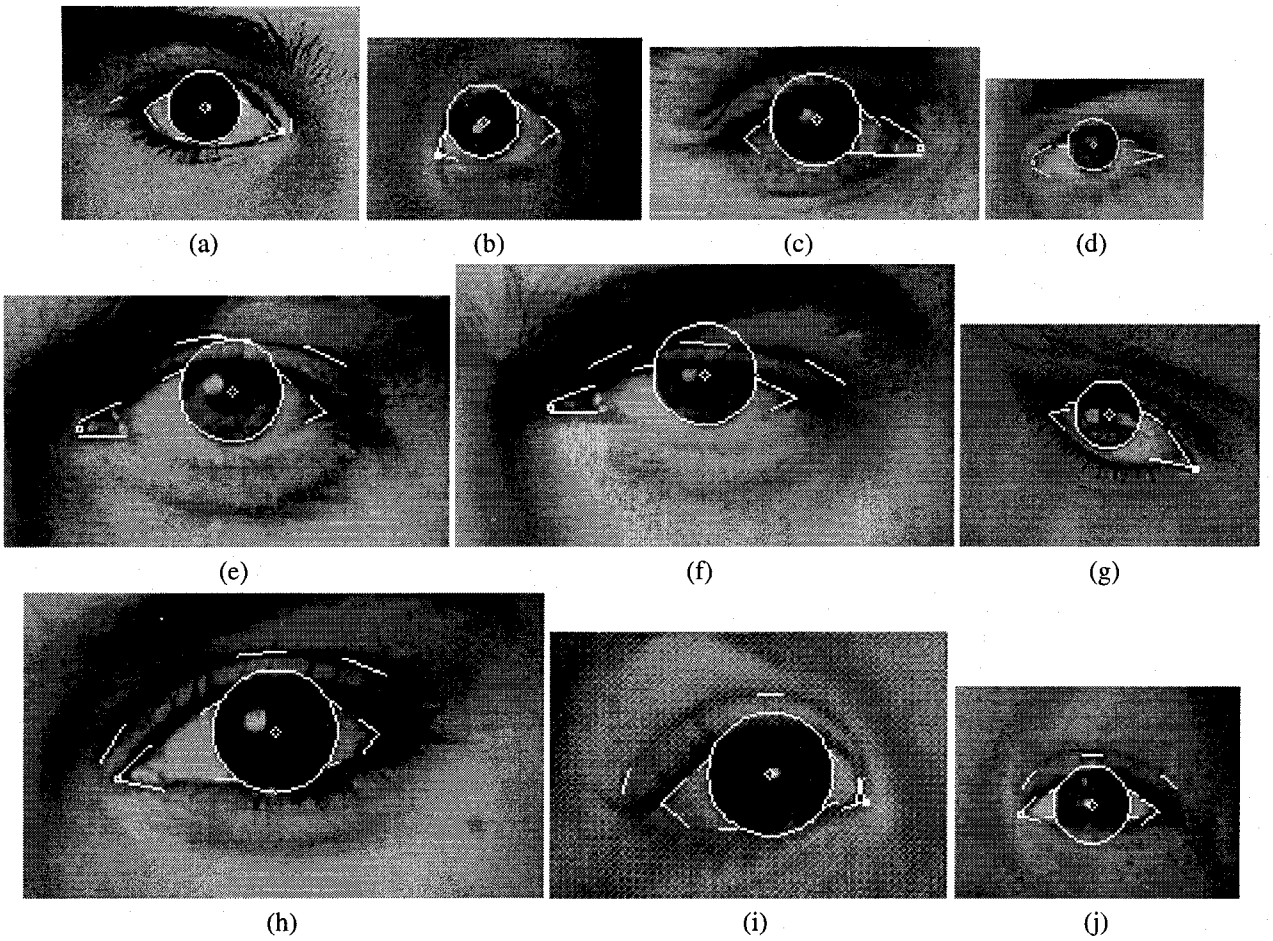


Figure 7. Ten examples of successfully analyzed eye regions. The different sizes of the eyes reflect their relative scale. A representative couple of processed images is shown.

advantage, that all the conditions, for example, that the iris has the shape of a circle and a minimum and maximum size, can be checked explicitly.

5. Results

The method has been tested on more than 100 face images, that means we have investigated more than 200 distinct eye and eyebrow regions. The several facial regions are localized and selected applying an attentive localization strategy [2]. Our face data base consists of normal faces[1] (volunteers of our institute) as well as of dysmorphic ones (provided from the Kinderzentrum München). The large variability of successfully processed facial regions is presented in figs. 7, 9, 11 and 12 . All the regions shown are processed applying a region dependent processing scheme all composed of the basic filter operations introduced before. A sample of representative examples are depicted in this contribution, which are not only the best results.

Eyes The overall error rate for the detection of the iris in eye regions was 2.3%. In 94.4% of the remaining eye regions the outer eye corner was detected successfully while the inner eye corner was correctly detected in 91.6% of the remaining cases. The reason for the higher error rate of the detection of the inner eye corner is its more complex structure and higher variability. The detection may fail if there are unexpected structures in the eye, e.g. a contact lens (fig. 8 left). If the resolution is to low and no discrimination between the border of the iris and the eye corner is possible the sequential search strategy will also fail (fig. 8 right). The same failure will occur if one wants to compute gravely squinted eyes.

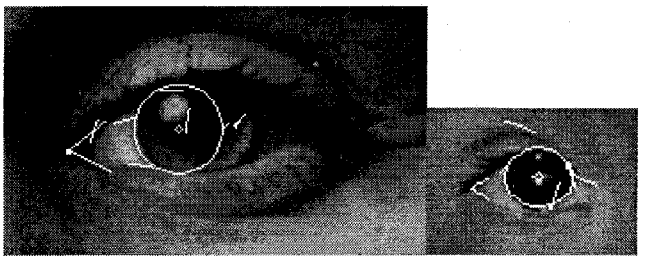


Figure 8. Examples at which the sequential search strategy fails. Caused by a contact lens (left) and by a too low resolution (right).

[1] some face images are taken from Manchester Face Database available via http://peipa.essex.ac.uk/ftp/ipa/pix/faces

Eyes of children with Sotos syndrome are shown in fig. 7b, 7i, and 7j. In these eyes the wrinkle of the upper eyelid is enlarged as far as the lower eyelid and covers the tear gland. Therefore, the tear gland is not visible in the inner eye corner. This feature is one of the most prominent feature or a leading dysmorphic sign of the sotos syndrome and it is very important for a detailed dysmorphic diagnosis. One example of eye region which is not in a horizontal orientation is depicted in fig. 7g. The developed algorithms can handle rotation angles up to 20°. Eye regions of different sizes are shown in fig. 7b (iris diameter = 28 pixel) and in fig. 7i (iris diameter = 53 pixel). In general, the images of Pentland's face database[2] are not in accordance with the necessary requirements (too low resolution) but to demonstrate the robustness of our methods some examples of eye regions are processed and one is depicted in fig. 7d (left eye of stephen).

Eyebrow The determination of the upper edge of the eyebrows is important for the calculation of the midface height (see also fig. 2). For the detection of the upper edge of the eyebrows we use the basic filter operation BFO1 in horizontal orientation with a large scale because large edge detection filters are more robust against fine structures (on a fine scale eyebrows have often no straight edges). The use of large scale filters is advantageous because they cover a larger part of the eyebrow edge and therefore, the filter response marks clearly the beginning of the edge (see fig. 9). We achieve an overall error rate of only 2.6% for the detection of the upper edge of the eyebrows in eyebrow regions.

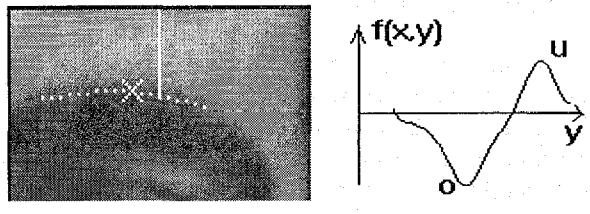


Figure 9. Detection of the upper eyebrow edge and the corresponding filter responses $f(x, y)$, computed along the white, vertical line. After the initial detection, the edge is tracked in both directions. The minimum (o) of the filter responses indicates the position of the upper eyebrow contour, while the maximum (u) that of the lower edge.

Mouth The detection of the mouth features are important for the determination of the width of the mouth and also for the determination of the midface height (see also fig. 2). The sequential search strategy developed for the detection of the mouth features is summarized in fig. 10. The overall error rate for the detection of the mouth ends is relatively worse with 16.2%. One main reason for the incorrect detection results is that an exact low-level definition of the mouth corner is extremly problematical (fig. 11). In our case we have defined the mouth corners related to the anatomical definition as that point at which the mucouses of the upper lip and of the lower lip (red lip) end. But often this position is not visible in the data and the detection is additionally complicated by edges of small ongoing mouth wrinkles or by shadows caused by small dimples near the mouth corners (fig. 11).

[2]available at anonymous ftp://whitechapel.media.mit.edu/pub/images/

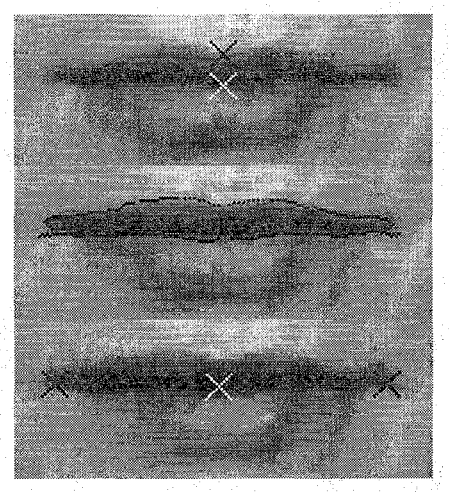


Figure 10. Sequential search strategy for the detection of the mouth ends. Initial detection of the upper edge of the upper mouth lip and the mouth fissure (first). The upper edge of the upper mouth lip and the lower edge of the mouth fissure are tracked in both directions until they end (second). The mouth corners are determined and the middle of the mouth is calculated (third).

Nose For the determination of the width of the nose an approximately horizontal edge caused by the nose wrinkles is detected and tracked at both sides of the nose. The overall error rate for the determination of the width of the nose was 5.6% (fig. 12). The detection will fail even if no sufficient edge information is remaining at one side of the nose (see fig. 12 lower right). The detection succeeds, however, also in cases of shadows if the remaining edge information is sufficient for the edge tracking (fig. 12 lower left).

6. Discussion

It has been shown, that by applying a sequential search strategy, complex facial keypoints can be detected very accurately. The proposed adaptive processing scheme provides a robust and fast method to detect and exactly localize keypoints in face images. For the detection of keypoints the global information of the whole region as well as the local structure of the keypoints are exploited. Only those keypoints are detected by our method, for which evidence

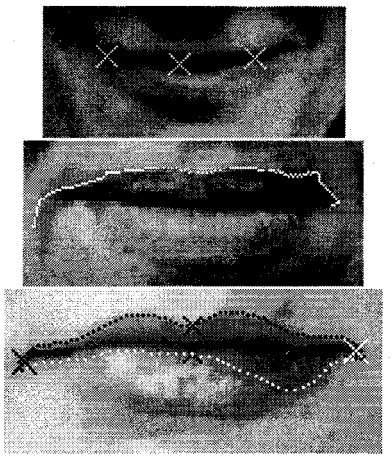


Figure 11. The detection of the mouth ends fails or may be inaccurately in cases of facial play (first), if shadows of dimples disturb the region (second), and if the tracking is misled by unexpected edges here caused by non frontal illumination (third).

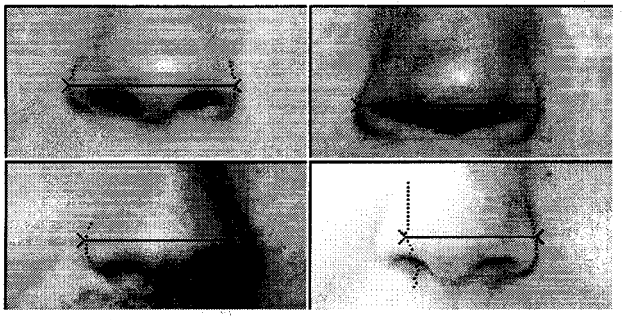


Figure 12. Examples of correctly determined widths of the nose (first row). Even in cases of shadows the detection succeeds (second row left). The detection fails if the remaining edge information is too low (second row right).

exists in the local structure of lines and edges. No extrapolations on the basis of other features are performed. If no detailed information of the local structure remains, the detection of the keypoint will fail (figs. 8, 11, and 12 lower right).

The robustness of the developed strategy origins from the use of model knowledge. At each step of the sequential strategy the model is used to control the next action. Furthermore, at each step the derived information can be checked for its consistency with the model.

In contrast to the approach of deformable templates [11] the sequential search strategy is more flexible concerning the adaptation to the local image structure. Our models consist of knowledge about the edges and lines, their orientation, length, scale, curvature, relative position, and vertices which have to be detected. However, only those details that are necessary are included and if possible only in a qualitative way and not with quantitative limits. An estimation of the position of not visible keypoints (in the data) as it is possible applying deformable templates is not desired, because estimated keypoint positions are not of interest for our application.

The time for the processing of one facial region depends mainly on the details considered during the processing. In practice, we convolve a facial region of a typical size of 150×100 pixels by 23 basis functions of a kernel size of 27×27 pixels what takes about 4 seconds on a 150 MHz R4400 SGI-Indy workstation. If the convolutions are done, the computation of the whole sequential search for an eye needs about half a second. Because of the one dimensional character of the sequential search (most of the time is spent for the line and the edge tracking) the computation time increases approximately linear with the size of the image.

Acknowledgments

We appreciate the support of Prof. Dr. S. Stengel–Rutkowski, of the Kinderzentrum München, Institute for Social Paediatrics and Youth Medicine. We are also grateful to our graduate student L. Witta.

References

[1] W.T. Freeman et al., *The design and use of steerable filters for image analysis*, IEEE-Trans. PAMI, Vol. 13, 891-906, 1991.

[2] R. Herpers et al., *GAZE: An attentive processing strategy to detect and analyze the prominent facial regions*, Proc. IWAFGR95, Zurich, M. Bichsel (Ed.) Switzerland, 214-220, 1995.

[3] R. Herpers et al., *Detection of keypoints in face images*, GSF-Bericht 23/95, GSF-Forschungszentrum, Germany, 1995.

[4] M.S. Kamel et al., *Face recognition using perspective invariant features*, Pat. Rec. Let., Vol. 15, 877-883, 1994.

[5] Y. Kaya et al., *A basic study on human face recognition*, in S. Watanabe (Ed.), Frontiers of Pattern Rec., Academic Press, 265-289, 1972.

[6] M. Michaelis, *Low level image processing using steerable filters*, PhD thesis, Universität Kiel, Germany, 1995.

[7] P. Perona, *Steerable-scalable kernels for edge detection and junction analysis*, ECCV'92, G. Sandini (Ed.), LNCS 588, Springer-Verlag, 3-18, 1992.

[8] P. Schimanek, *Anthropologische und anthropometrische Definition von Dysmorphiezeichen als Grundlage der Diagnostik von Dysmorphiesyndromen*, PhD thesis, Universität München, Germany, 1988.

[9] S. Stengel-Rutkowski et al., *Anthropometric definitions of dysmorphic facial signs*, Hum. Genet, vol.67, 272-295, 1984.

[10] S. Stengel-Rutkowski et al., *Chromosomale und nicht-chromosomale Dysmorphiesyndrome*, Enke Verlag, Stuttgart, 1985.

[11] A.L. Yuille et al., *Feature extraction from faces using deformable templates*, Proc. of IEEE Conf. CVPR '89, 104-109, 1989.