Context Based Detection of Keypoints and Features in Eye Regions*

Rainer Herpers Markus Michaelis
Gerald Sommer[1] Lars Witta[2]

GSF – Institut für Medizinische Informatik und Systemforschung, Medis, Ingolstädter Landstr. 1,
85764 Oberschleißheim, Germany, Email: (herpers, michaeli)@gsf.de
[1] Institut für Informatik, Christian-Albrechts-Universität, Preußerstr. 1-9, 24105 Kiel, Germany
[2] current address: Lehrstuhl für Mensch-Maschine-Kommunikation, TU-München, Arcisstr. 12, 80290 München, Germany

Abstract

Facial keypoints such as eye corners are important features for a number of different tasks in automatic face processing. The problem is that facial keypoints rather have an anatomical high-level definition than a low-level one. Therefore, they cannot be detected reliably by purely data-driven methods like corner detectors that are only based on the image data of the local neighborhood.

In this contribution we introduce a method for the automatic detection of facial keypoints. The method integrates model knowledge to guarantee a consistent interpretation of the abundance of local features. The detection is based on a selective search and sequential tracking of edges controlled by model knowledge. For this, the edge detection has to be very 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 are important features for a number of different tasks in automatic face processing [4, 5]. In this contribution we will introduce a method for the automatic detection and localization of such keypoints. The problem is that the keypoints are defined rather as anatomical features (e.g. the corner of an eye) than by a low-level definition like a junction. Different examples of the same facial keypoints can be very different in terms of their grey value distribution in the image (fig. 1). Hence, it is not possible to detect facial keypoints by standard corner detectors [6] or other purely local and data-driven detectors [10] that do not make use of context or 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. Yuille et al. [13, p.104] have written: "*The problem seems to be that although it is straightforward to find local evidence for edges, it is hard to organize this local information into a sensible global percept*". The integration of appropriate model knowledge is mandatory. We believe that this situation is typical for complex real world problems. In Yuille et al. deformable templates are used to obtain such a global percept. In the case of the eyes, for example, these templates consist of two parabolic sections for the eyelid edges and of a circle for the iris. Applying deformable templates the number of parameters to tune the model to the data is limited. Therefore, an exact localization of the eye corners according to the anatomical definition of the keypoints is not guaranteed because the template is not flexible enough to adapt to all details.

Figure 1. Three examples of inner eye corners. The large variance of eye corners does not allow for a purely local and data-driven detection.

Our approach starts by detecting the most prominent and reliable features in the eye region. Given our task and image recording conditions these are the strong vertical step edges of the iris. Subsequently, the complete iris and the eyelids are tracked to finally detect the eye corners. At each step the detection and tracking is controlled by the integrated model knowledge. The already detected edges are checked for their consistency as well as the specific edge structures which are searched for in the next step are given by the model.

This work is partially supported by DFG grants So 320/1-1 and Ei 322/1-1.

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

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 (fig. 2, [12, p. 42]). The ratios of distances between certain facial keypoints are statistically significant for discriminating between normal children faces and different classes of dysmorphic syndromes [11, 12]. Therefore, the detection of particular keypoint positions in dysmorphic face images is of a high diagnostic value (see fig. 7, [11, p. 63]). For this, the localization of the keypoints should be precise, reproducible, and it should correspond to the anatomical definition of the keypoint positions.

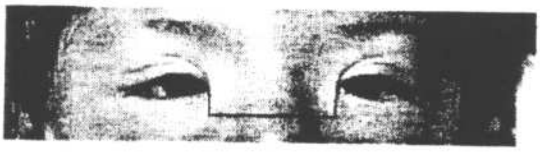


Figure 2. Example of a very enlarged intercanthal distance which is a typical dysmorphic facial sign (taken from [12, p. 42]).

The organization of the paper is as follows. The next section gives a brief outline of the filtering scheme that we use for the edge detection and tracking. Subsequently we will introduce the sequential search strategy for the example of an eye region. In section 4 we will present results followed by a short discussion.

2. The filtering scheme

The filtering scheme is essentially based on an edge detector. We use a first derivative of Gaussian:

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

$C(\sigma)$ is a normalization constant such that the filter has a L^1 norm of 1 for all scales. For the aspect ratio ε there are the following considerations. Elongated filters have proven to show good results for edge and line detection [9, 8]. However, very elongated filters are computationally expensive if they are steered in orientation. In addition, it is relatively easy to synthesize more elongated filters from less elongated ones but the opposite is more difficult. As a compromise, we choose an edge detection filter with an aspect ratio of 2 (fig. 3). For some tasks we also use a line detector based on second derivative of Gaussian filter.

Steerable filters were introduced to efficiently calculate the responses of filters in arbitrary orientations, scales, and other deformations [1, 9]. It refers to the reconstruction of all deformed filters F_α by a superposition formula of the following type:

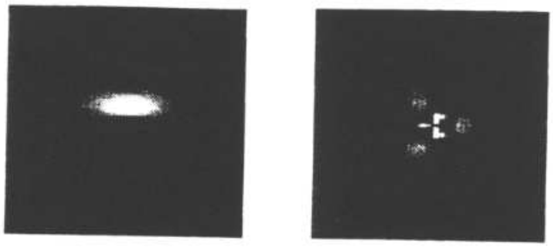


Figure 3. The mother filter (left) and a typical basis function (right) to steer the filter in orientation and scale.

$$F_\alpha(\vec{x}) = \sum_{k=1}^{N} b_k(\alpha)\, A_k(\vec{x}) \qquad (2)$$

α is an arbitrary deformation (multi-) parameter. In our case the orientation and scale are steered. In contrast to the work of Freeman & Adelson [1] the term 'steerability' here is applied to all deformations, not just to rotations. The number N of so called **basis functions** $A_k, k = 1 \ldots N$ is small compared to the number of deformed filters. Typically N will be 10 or 20, while α theoretically assumes an infinite number of values and many thousands in practice for orientation and scale.

The quality of the reconstructed filters depends on the number of basis functions (bfcts). Figure 4 shows examples of reconstructions with different numbers of basis functions. We don't use deformed copies of the filter as basis functions but orthogonal basis functions instead (fig. 3 right). For more theoretical background we refer to [7]. The following two properties of orthogonal basis functions are essential for the design of our filtering scheme. 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 filters are sufficient because they qualitatively still resemble elongated edge detectors (fig. 4 left). Therefore, the region to be analyzed is convolved with only a small number of basis functions. More basis functions are added only during the processing at certain positions where better approximations are required.

The detection of the eye corners is based on a sequential search and tracking of the edges in the eye region. The detection and tracking is performed by three different **basic filter operations** that make extensive use of the steered filter.

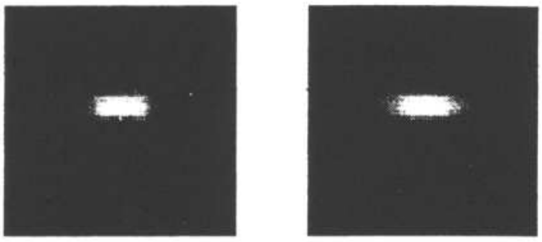
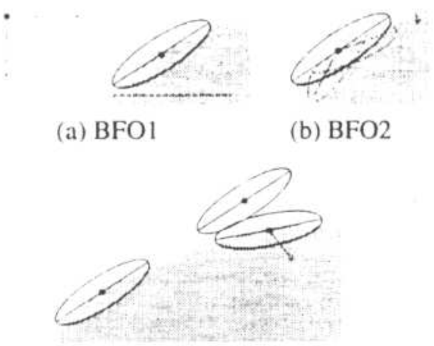


Figure 4. Reconstruction of the filter with different numbers of basis functions. 10 bfcts. (left) and 30 bfcts. (right) are used for steering orientation and scale (2 octaves). The respective L^2 errors are 22% and 3%.

- The first operation (BFO1) searches in a predefined region for an edge with a predefined orientation and scale (fig. 5a).
- The second operation (BFO2) determines the orientation of an edge at a given position by evaluating the maximal response of a rotated filter (fig. 5b).
- The third operation (BFO3) tracks an edge by one step. For this, the filter is moved by a small step in the already known direction. Then the edge is searched again in perpendicular direction (fig. 5c).



(a) BFO1 (b) BFO2

(c) BFO3

Figure 5. Basis filter operations. BFO1: Detection of an edge (a). BFO2: Determination of the orientation (b). BFO3: Stepwise tracking of an edge (c).

3. Sequential search strategy

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 for a full face images), the orientation (frontal, not tilted), the illumination (frontal and diffuse), as well as the restriction to faces without glasses or other occlusions.

The detection of the keypoints is achieved by a sequential search respectively tracking of the edges within the eye region where each step consists of several applications of the basic filter operations. The selection of the different operations and their parameters for each step is controlled by the already derived information together with the model knowledge. The model knowledge used consists of relevant edges, their scales and geometrical relations of the considered object respectively of an eye region as depicted in fig. 6. This information is used to search for reliable edges at specific positions, orientations, and scales in each step of the sequential search. Furthermore, different kinds of edges (white-to-black against black-to-white edges or edges against lines etc.) can be distinguished. A complete description of all details of the model used is beyond the scope of this paper, but it can be found in [3].

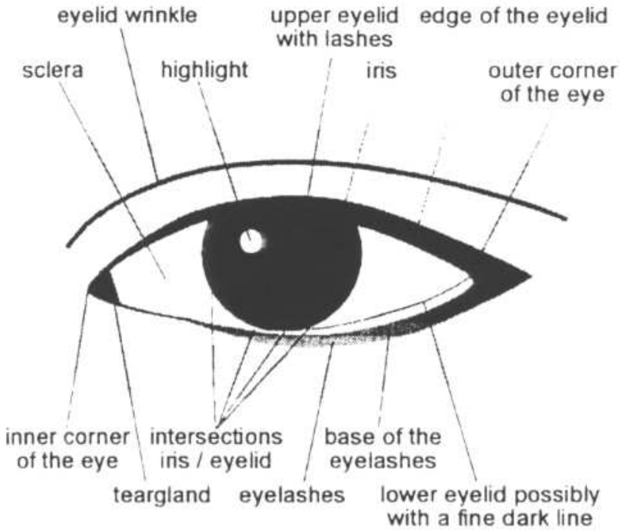


Figure 6. Model of the left eye. The depicted details are important to control the sequential search.

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

To better understand the sequential search strategy, the detection of the iris and of the eye corners will be described in more detail (fig. 8). The most prominent and reliable features within the eye region are the edges of the iris. Therefore, the sequential search starts by detecting these edges. The iris is not detected in one step by applying a especially designed circular symmetry filter because it is faster and more reliable to divide the detection into several steps. First

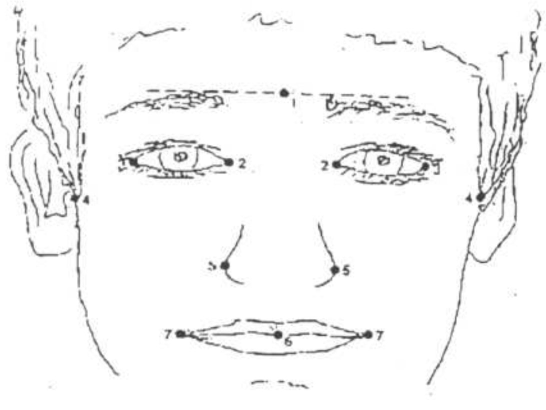


Figure 7. Keypoints in a frontal face image. In this contribution only the points number 2 and 3 (inner and outer eye corners) are of interest. The figure is taken from [11, p. 63].

the left edge segment (vertical, white-to-black edge) of the iris is searched for in the selected image part applying the basis filter operation BFO1 (fig. 5a). The spatially limited image parts of the eye regions considered are computed by applying an attentive localization algorithm [2]. The detected edge is tracked upwards and downwards (using BFO3) until the intersections 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. 8 first row right). The computed features are checked with the model knowledge and if they do not agree with the expected circular symmetry, size of the radius, and other constraints, the search starts again to detect the next prominent vertical edge segment. By applying this processing, the detection of a highlight often visible on the iris can be excluded.

After the detection of the intersection points of the iris and the eyelids (fig. 8 second row), the edges of the eyelids are tracked. The tracking is controlled at each step by the model knowledge to avoid e.g. that the tracking is misleaded by edges of the tear gland. This processing differs strongly from a 'blind' edge detection step with a subsequent interpretation of the edge image. Only edges with properties known by the model are considered. Furthermore, for edges, which are expected from the model, our search algorithm is looking explicitly for hints in the image data. Therefore, a robust and reliable edge detection and tracking is achieved because only edges, that are in accordance with the model (especially their orientation), are considered during the processing. The application of orientation selective filters enhances the robustness of the edge detection and tracking compared to the application of isotropic edge detectors if there are competing edges with different orientations.

Finally, the inner eye corner is detected at that point on the curved edge segment, which has the greatest distance from the center of the iris (fig. 8 last row left). The processing of a left and of a right eye is distinguished by the search algorithm. The search strategy is slightly different for inner and outer eye corners because of the missing tear gland and other details (fig. 8 last row).

The proposed method can deal with a high degree of variability of the image data of the objects to be considered (in the presented case the eye region) (fig. 9). Predefined fixed distances, angles or scales are replaced as far as possible by variable ranges with relative limits, which are calculated from model parameters or actually acquired knowledge during the runtime of the sequential search algorithm. This enhances the robustness and efficiency of the developed algorithm.

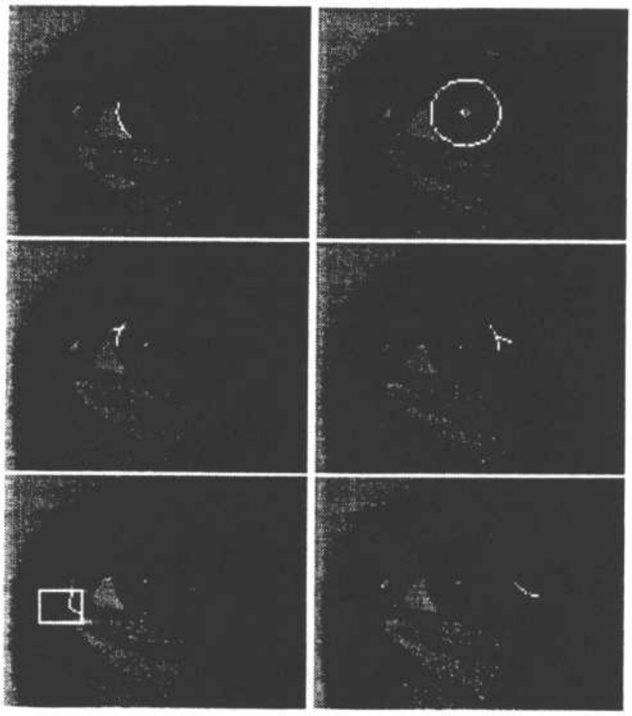


Figure 8. Sequential search strategy depicted for a left eye. First, a prominent vertical white-to-black edge is detected. After the detection of the corresponding right edge segment, the final segmentation of the iris is computed (first row). Detection of the intersection points of the edge segments of the iris with the edges of the upper and the lower eyelid (second row). The eyelid edges are tracked and finally an 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 (third row).

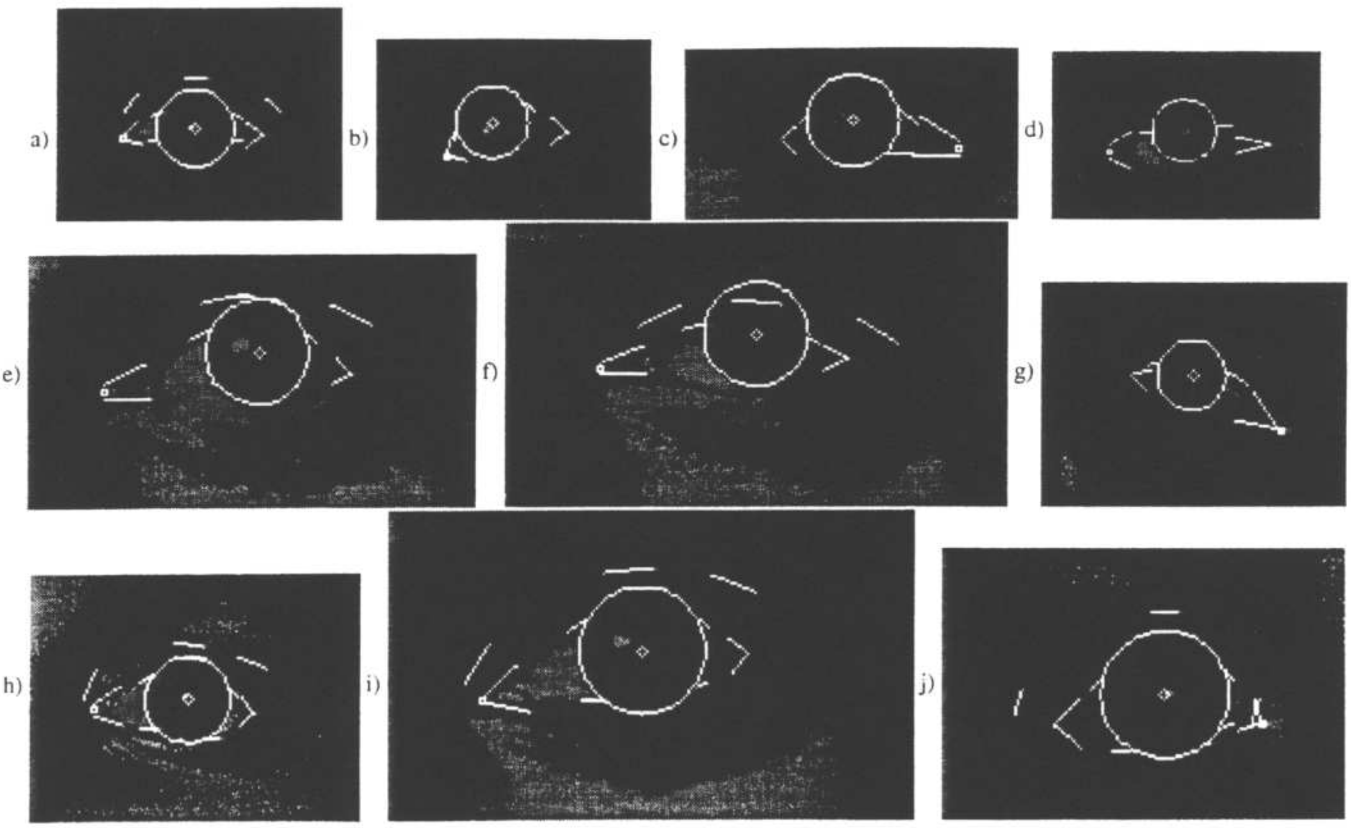


Figure 9. Ten examples of successfully analyzed eye regions. The different sizes of the eyes reflect their relative scale.

4. Results

The method has been tested on eye regions from more than 100 different face images. The data base consists of normal faces as well as of dysmorphic faces. The overall error rate for the detection of the iris was 2.35%. In 94.4% of the remaining eye regions the outer eye corner was detected successfully while the inner eye corner was detected in 91.6% correctly. The reason for the higher error rate of the detection of the inner eye corner is its more complex structure and higher variability (fig. 1) but the sequential search algorithm presented can easily be improved by additional consistency checks, so that the error rate will decrease. The large variability of successfully processed eye regions is demonstrated in fig. 9. All the depicted eye regions are processed applying the same model and parameter settings.

Eyes of children with Sotos syndrome are shown in fig. 9a, 9b, and 9j. In two cases the wrinkle of the upper eyelid is enlarged as far as the lower eyelid so that it covers the tear gland. Therefore, the tear gland is not visible in the inner eye corner. This feature is one of the characteristic dysmorphic signs for the sotos syndrome and it is important for a detailed dysmorphic classification.

One eye region which is not in a horizontal orientation is depicted in fig. 9g. Rotation angles up to 20 degree are tolerated without explicitly changing the parameter settings of the sequential search algorithms.

Eye regions of different sizes are shown in fig. 9g (iris diameter = 28 pixel) and in fig. 9j (iris diameter = 54 pixel). The diameter of the iris varies between 14 pixel and 104 pixel (not depicted) for the largest successfully investigated eye image in our database. At all scales except for the smallest one, the detection of the iris as well as of the eye corners is possible. At the smallest scale the detection of the eye corners fails because of missing details, which are assumed by the model, to guide the line and edge tracking.

To demonstrate the robustness of the presented method, different examples of eye regions selected from the face database of A. Pentland[1] have been processed. One of them is depicted in fig. 9d (left eye of stephen) where the detection is successful although the images of the database are too low in resolution in general.

The developed algorithms are robust against noise. We added noise with a SNR between 14dB and 4.5dB. In any case the iris can be detected successfully, but the exact detection of the eye corners may fail if the SNR is too low.

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

5. Discussion and conclusion

It has been shown, that by applying a sequential search strategy, complex facial keypoints can be detected very precisely. The proposed adaptive processing scheme provides a robust and fast method to exactly localize keypoints in face images.

For the detection of keypoints the local structure at its location but also the 'global' edge information in the near surrounding is exploited. The robustness is achieved by many global and local consistency checks between the object and the model. Hence the included model knowledge is able to actively guide the search, i.e. the tracking of the edges, based on the geometrical constraints of the expected edges in the eye region.

Only those keypoints are detected by our method, for which evidence exists in the local structure of lines and edges. No extrapolations on the basis of other features are performed.

We want to point out, that the specific eye model presented does not impose any restrictions in the flexibility of our method. It merely represents the knowledge about the class of images we have to consider.

In the case of large variability of the image data to be considered the sequential search strategy is more adaptive in contrast to the deformable template approach [13]. Therefore, it provides more reliable localizations of the keypoints.

The sequential processing strategy is based on a powerful filtering scheme which provides all the features needed for a robust detection and tracking of the edges. The tradeoff between the quality and the speed of the filters can easily be adjusted on-line by varying the number of basis functions. For example, only two basis functions are necessary for the detection of the prominent vertical step edge at the beginning of the sequential search strategy (fig. 8 first row), because the full orientation selectivity of the filter is not required. For all different search and tracking tasks a maximum of 10 basis functions for the edge detection is sufficient. Also the choice of orientation selective edge detectors enhances the robustness of the edge tracking compared to less orientation selective ones. This is especially important if there are junctions with misleading edges.

The time for the processing of one eye region depends mainly on the details to be considered. In practice, we convolve an eye region of a typical size of 120×70 pixel by 10 basis functions of a kernel size of 27×27 pixel what takes about 3 seconds on a 150 MHz R4400 Silicon Graphics Indy workstation. If the convolutions are done, the computation of the whole sequential search shown in fig. 8 needs about half a second. Because of the one dimensional character of the sequential search (most of the time is spent for the edge tracking) the computation time increases approximately linear with the size of the image.

A precise localization of the prominent facial keypoints and features is very important for many problems in face recognition and face identification. This contribution demonstrates a successful example for an efficient solution of a complex detection problem in real world images. Presently the method is extended to detect other facial keypoints, such as the corners of the mouth or the outline of the nose.

Acknowledgments

We appreciate the support of Prof. Dr. S. Stengel–Rutkowski, of the Kinderzentrum Munich and Institute for Social Paediatrics and Youth Medicine. We are also grateful to our graduate students H. Kattner and H. Rodax.

References

[1] W. Freeman and E. Adelson. The design and use of steerable filters for image analysis. *IEEE Trans. PAMI*, 13:891–906, 1991.

[2] R. Herpers, H. Kattner, H. Rodax, and G. Sommer. GAZE: An attentive processing strategy to detect and analyze the prominent facial regions. In *Proc. of International workshop on face– and gesture recognition, IWAFGR95, Zurich, M. Bichsel (ed.)*, pages 214–220, 1995.

[3] R. Herpers, M. Michaelis, L. Witta, and G. Sommer. *Detection of keypoints in face images*. Tech. Report 24/95, GSF–Research Center, Germany, 1995.

[4] M. Kamel et al. Face recognition using perspective invariant features. *Pattern Recognition Letters*, pages 877–883, 1994.

[5] Y. Kaya and K. Kobayashi. A basic study on human face recognition. In *Frontiers of Pattern Recognition, S. Watanabe (ed.), Academic Press*, pages 265–289, 1972.

[6] L. Kitchen and A. Rosenfeld. Gray-level corner detection. *Pattern Recognition Letters*, 1:95–102, 1982.

[7] M. Michaelis. *Low Level Image Processing Using Steerable Filters*. PhD Thesis Christian-Albrechts-Universität, Kiel, 1995.

[8] M. Michaelis and G. Sommer. Junction classification by multiple orientation detection. In *Proc. of ECCV 94, Springer, LNCS 800, Vol.I, J.–O. Eklundh (ed.)*, pages 101–108, 1994.

[9] P. Perona. Steerable-scalable kernels for edge detection and junction analysis. In *Proc. of ECCV 92, Springer, LNCS 588, G. Sandini (ed.)*, pages 3–18, 1992.

[10] L. Rosenthaler, F. Heitger, O. Kübler, and R. von der Heydt. Detection of general edges and keypoints. In *Proc. of ECCV 92, Springer, LNCS 588, G. Sandini (ed.)*, pages 78–86, 1992.

[11] P. Schimanek. *Anthropologische und anthropometrische Definiton von Dysmorphiezeichen als Grundlage der Diagnostik von Dysmorphiesyndromen*. PhD Thesis Universität München, Germany, 1988.

[12] S. Stengel-Rutkowski and P. Schimanek. *Chromosomale und nicht chromosomale Dysmorphiesyndrome*. Enke Verlag, Stuttgart, Germany, 1985.

[13] A. Yuille, D. Cohen, and P. Hallinan. Feature extraction from faces using deformable templates. In *Proc. of IEEE Conf. CVPR '89*, pages 104–109, 1989.