HIERARCHICAL FILTERING SCHEME FOR THE DETECTION OF FACIAL KEYPOINTS

Markus Michaelis[1] *Rainer Herpers*[1] *Lars Witta*[2] *Gerald Sommer*[3]

[1] GSF – Institute of Medical Informatics and Health Services Research, MEDIS
D-85764 Neuherberg, Germany michaeli@gsf.de
[2] Lehrstuhl für Mensch-Maschine-Kommunikation, Technical University, D-80290 München, Germany
[3] Computer Science Institute, Christian-Albrechts-University, D-24105 Kiel, Germany

ABSTRACT

Usually, the first processing step in computer vision systems consists of a spatial convolution with only a few simple filters. Therefore, information is lost or it is not represented explicitly for the following processing steps. This paper proposes a new hierarchical filter scheme that can efficiently synthesize the responses for a large number of specific filters. The scheme is based on steerable filters. It also allows for an efficient on-line adjustment of the trade off between the speed and the accuracy of the filters. We apply this method to the detection of facial keypoints, especially the eye corners. These anatomically defined keypoints exhibit a large variability in their corresponding image structures so that a flexible low level feature extraction is required.

1. INTRODUCTION

Face recognition is of importance for many tasks like man-machine-communication, surveillance, and others. Many methods in face recognition need the detection of facial keypoints like the eye-corners [5]. The problem in detecting these keypoints is the complicated structure and the extremely large variability of the corresponding image structures. Therefore, purely data-driven corner detectors like the one of Kitchen & Rosenfeld [6] are not able to detect these keypoints. Yuille et al. proposed a deformable template approach to incorporate a 'global percept' of the scene [9]. However, due to the small number of parameters in their models the exact localization of the keypoints is not guaranteed. To cope with this situation in this paper a powerful low level feature extraction scheme is suggested.

Usually, the first processing step in computer vision systems consists of a spatial convolution with only a few simple filters. This restriction, however, has the drawback that information is lost or that it is not represented in an explicit way. Hence, the following processing steps are not optimally supported. To obtain more complete descriptions, more flexible filtering schemes are required.

Recently a new filtering technique called steerable filters (other names have been used also) has been proposed to obtain the response of a certain mother kernel in a continuum of orientations, scales or other parameters [1, 8]. The idea of this technique is to apply a small number of basic kernels, so called basis functions, which have been chosen appropriately so that all kernels of interest can be generated from them by superpositions. In previous work [1, 8] only 'primitive' edge or line detection kernels have been steered by this method. In contrast to this, we are interested in better exploiting the large flexibility offered by the steerability scheme to synthesize a variety of more complex kernels, e.g. a kernel with circular symmetry. For this we propose a new hierarchical filtering scheme. Moreover, we propose to vary on-line the number of basis functions to optimize the trade off between the speed and the accuracy of the filters.

This work is partially supported by the DFG grants So 320/2-1 and Ei 322/2-1.

2. STEERABLE FILTERS

The term 'steerable filters' refers to the reconstruction of deformed filter kernels F_α (respectively their responses, α is a general multi-deformation) by a superposition formula of the following type:

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

The number N of so called **basis functions** $A_k, k = 1 \ldots N$ is assumed to be small compared to the number of deformed kernels. 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). Our method to obtain the optimal basis functions is based on Perona's approach [8]. One main difference is that we explicitly address the finite dimensional case which is the relevant for practical problems. Furthermore, we steer all deformations (e.g. orientation and scale) in one step. Only the basic idea of the method will be given. More theoretical background can be found in [7].

Let $F_k(x, y)$ denote the deformed kernels, where 'k' samples all deformations together (orientation and scale). $\langle \cdot | \cdot \rangle$ denotes the usual scalar product.

The matrix G with elements $G_{kl} = \langle F_k | F_l \rangle$ is called the Gramian. It is real and symmetric and therefore, it has a complete set of eigenvectors. The eigen decomposition of G is denoted as:

$$\langle F_k | F_l \rangle = \sum_m u_{m,k} \gamma_m u_{m,l} \qquad (2)$$

u_m denotes the m'th eigenvector with eigenvalue γ_m. $u_{m,k}$ is k'th element of m'th eigenvector.

The optimal (and orthogonal) basis functions $A_m(x, y)$ are given by

$$A_m = \sum_l u_{m,l} F_l \tag{3}$$

'Optimality' here means that a minimal number of basis functions is necessary for a given L^2 error. The interpolation functions for arbitrary orientations (θ), scales (α) etc. are given by

$$b_m(\theta, \alpha) = \frac{\langle F_{\theta,\alpha} | A_m \rangle}{\| A_m \|^2} \tag{4}$$

The sampling of θ and α in this formula is independent of the sampling for the Gramian. The reconstruction of the deformed kernels is by $F_{\theta,\alpha} = \sum_m b_m(\theta, \alpha) A_m$. It should be emphasized that the calculation of the b_m and A_m by (2), (3), and (4) is done off-line and thus it is not time critical. Examples of basis functions to steer a Gaussian first derivative kernel (an edge detection filter) in orientation and scale are depicted in fig. 2 (top box).

2.1. Properties of the basis functions

The quality of the reconstructed kernels depends on the number of basis functions. In previous work a fixed number of basis functions has been used for the reconstruction of the kernels. We point to the benefits of orthogonal basis functions for the on-line adaptation of the quality and speed of the filters by changing the number of basis functions. Figure 1 shows examples of reconstructions with different numbers of basis functions. The following two properties of the basis functions are essential:

- 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 kernels. Only the quality of the reconstruction changes.

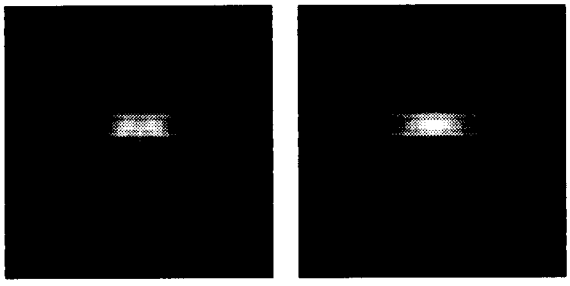

Figure 1. An elongated first Gaussian derivative kernel reconstructed with different numbers of basis functions. Depicted are reconstructions of the kernel with 10 (left) and 30 (right) basis functions. The respective L^2 errors are 22% and 3%.

In most cases low quality approximations of the kernels are sufficient. Therefore, the region that is to be analyzed is convolved with only a small number of basis functions. The reconstructions then have a relatively large error. Nevertheless, they qualitatively resemble the original kernel. More basis functions are added only during the processing at certain positions where better approximations are required.

3. HIERARCHICAL FILTERING SCHEME

In the first step of the hierarchical filtering scheme the image is convolved with a small number of basis functions (fig. 2, top box). These are the **only** convolutions that are really carried out. All following filter responses are generated by superpositions of these responses. In the next step an elongated first derivative of Gaussian kernel is steered in orientation and scale (other kernels and deformations are possible)(fig. 2 middle box). We will call these kernels the 'primitive' kernels. The aspect ratio is 2 to enhance the orientation selectivity of the kernel.

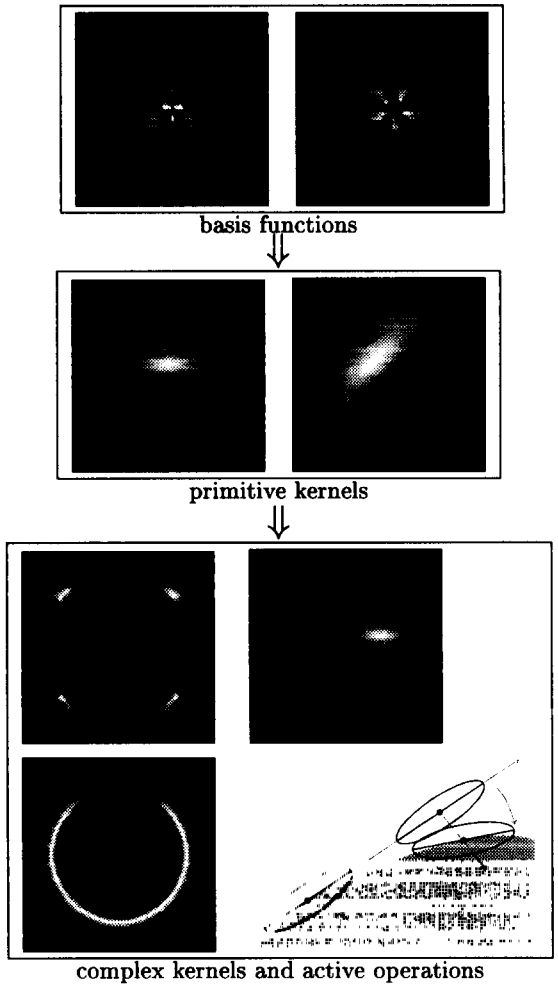

Figure 2. The hierarchical filter scheme. First step (top box): examples of basis functions. Second step (middle box): primitive kernels with different orientations and scales. Third step (bottom box): circular kernel with 4 primitive kernels and 10 basis functions for each (top left). Circular kernel with 28 primitive kernels and 30 basis functions for each (bottom left). The gap in the circular kernel is motivated by its use as an iris-detector. An one-sided kernel that is rotated around a shifted center (top right). An active edge tracking operation that searches the next edge element (bottom right).

For the third step we exploit the possibility to reconstruct

primitive kernels at arbitrary positions, scales and orientations. We can then synthesize more complex kernels from superpositions of the primitive kernels. Examples of such complex kernels are an one-sided and a circular kernel (fig. 2 third box). In addition to the complex kernels some basic 'active operations' can be derived from the steered primitive kernels. An example is an one-step edge-tracking operation (fig. 2 third box). For this the filter scans the vicinity of a given edge element in orientation and space to detect the next edge element.

The reuse of the same set of basis functions for many different kernels makes the scheme very flexible and efficient. The use of steerable filters as the basis for the complex filters and the edge tracking operation offers a tremendous flexibility for the feature extraction in low level vision systems.

4. DETECTION OF FACIAL KEYPOINTS

For the detection of facial keypoints we propose a model driven approach. A model, e.g. of the eye, is used to control the feature extraction and to guide a sequential search for the keypoints. These parts of the whole facial keypoint detection scheme have been already presented in other publications [2, 3]. In this paper we focus on the hierarchical filter scheme that is used to support a flexible and complete low level feature extraction. Appropriate complex filters are synthesized on-line based on primitive filters to derive the features of interest. Especially we present an improved iris detection approach and a final verification step for the eye corner candidates that are detected by the approach in [2, 3].

A basic idea of the suggested filter scheme is that simple but fast filters are applied first. If the responses of the fast filters are ambiguous or if more specific features are required more time consuming filters are applied. However, this will be the case only for certain positions or filter parameters (orientation etc.). The complex filters can be applied in simpler and faster versions first, by using less primitive filters for their synthesis. In addition, the number of basis functions for the reconstruction of the primitive filters can be varied (section 2.1.). This idea is demonstrated by the following examples.

The first example is the detection of the circular symmetry of the iris (fig. 3) which is the most reliable feature of the eye. Therefore, the search starts by detecting the iris with a fast circular filter (fig. 3, left). For this the image is convolved with 10 basis functions for steering an elongated Gaussian first derivative filter. If steering the scale of the primitive filters is not required, as in this special example, even less basis functions were necessary.

The fast circular filter has a spacing of 90° between the constituent primitive kernels. The response of a high-quality circular filter with 30 basis functions and 10° spacing (fig. 3, right) is calculated only at those positions, where the low-quality response is above a certain threshold (for demonstration purposes it is calculated here for all positions). Usually only a few percent of the eye region pixels have to be processed by the accurate but time consuming filter. The projections to the additional basis functions and primitive kernels are calculated only for these pixels. Hence,

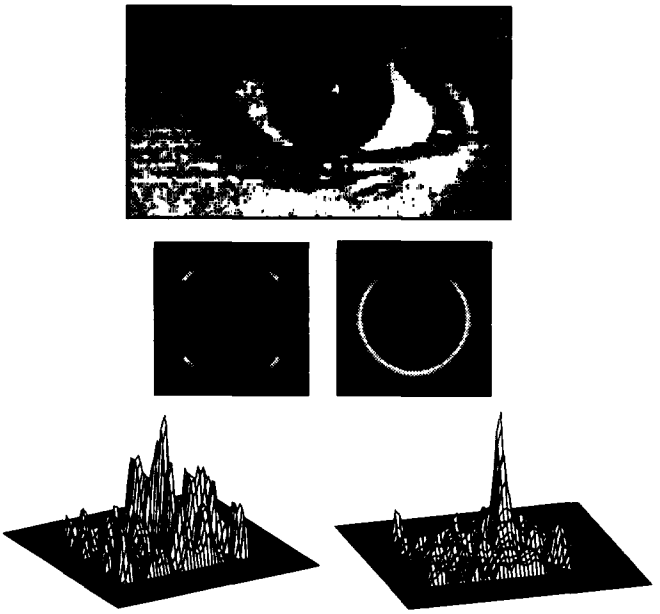

Figure 3. Top: Example eye region. Middle: A fast (left) and an accurate (right) version of a circular iris detector kernel. Bottom: Responses of the circular filters to the eye region as wire frame plots. The largest peak indicates the center of the iris.

we obtain the performance of the high-quality filter with the costs of the simple filter which is about 20 times faster.

The responses of fig. 3 are for a circular filter with an appropriate radius. If the radius of the iris is unknown, different radii of the iris have to be tested. Here again, the fast version of the circular filter allows a significant speed up without losing performance.

After the iris is detected the upper eye lid is tracked by the edge tracking operation of fig. 2 (see [2, 3] for more details). The position where the eye-lid changes its orientation is a candidate for the eye-corner. However, because of the complicated and variable image structures false detections of the eye corners may occur. Therefore, by applying an one-sided filter (fig. 2, bottom box) the candidate points are tested for the presence of a V-junction that is likely to be an eye corner (fig. 4).

The synthesis of the complex kernels is not perfect because the responses of the basis functions are available only on a discrete grid. For rotations and other deformations of the complex filters the primitive responses are also needed between pixels. The strategy is again to first apply a fast, non-interpolated one-sided filter. The interpolated filter is calculated only at those positions and orientations where more details are of interest, e.g. where the response is above a certain threshold.

The interpolated filter is derived by a spatial linear interpolation of the responses of the primitive filters at pixel positions. For a four nearest neighbor linear interpolation the filter is almost perfect but it is about 5 times slower than the non-interpolated one. However, the interpolated response has to be calculated only for those positions where the non-interpolated response is ambiguous. Figure 4 shows

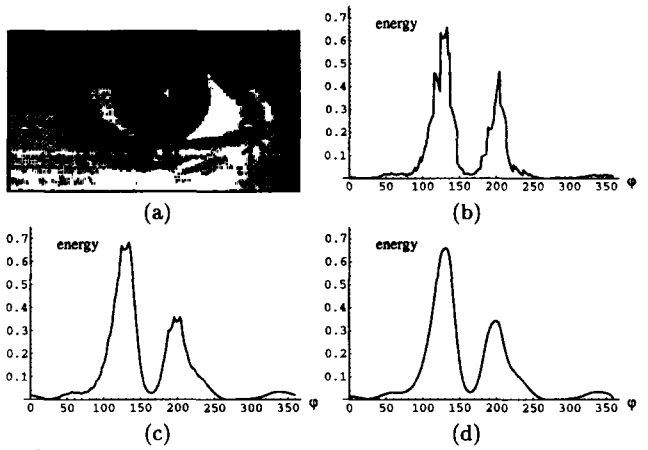

Figure 4. Orientational energy signature for the one-sided filter at the right eye corner (arrow) (a). Depicted are the responses of the non-interpolated filter (b), the 4-nearest-neighbor interpolated filter (c), and the smoothed response of the interpolated filter (d).

an example. The 'jaggedness' of the non-interpolated response (fig. 4b) is caused by the discretization of the position. With a subsequent smoothing, however, even the non-interpolated response gives good results. The interpolated and smoothed response has no visible difference to the response of the true one-sided filter.

Figure 5 shows examples of successfully analyzed eye regions. The large variability in the depicted eye regions demonstrates the robustness of the approach. Although there are many irritating structures from wrinkles, eyebrows, shadows etc. the anatomically correct keypoints are detected successfully.

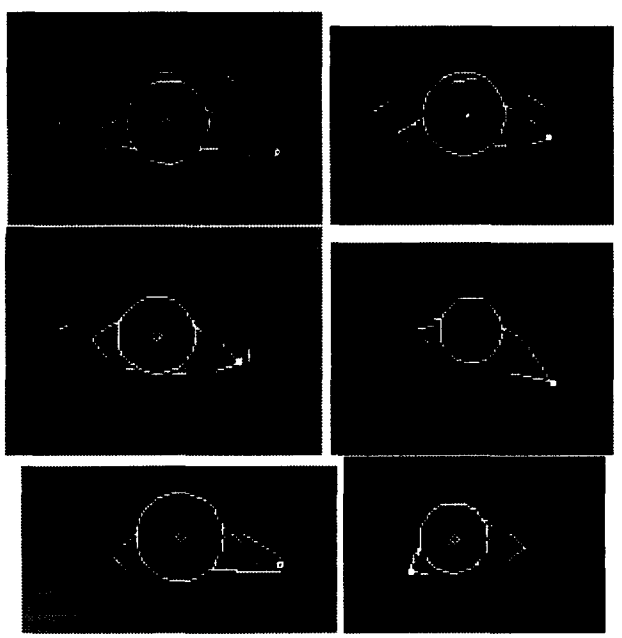

Figure 5. Examples of successfully analyzed eye regions.

5. CONCLUSIONS

In this paper we focused on a new hierarchical filter scheme that has been applied to improve a model based facial keypoint detection approach [2, 3]. The scheme offers a large flexibility for the feature extraction in low level vision systems. Model knowledge together with the information that is derived during the processing controls the choice of the filters. The trade off between the quality and the speed of the filters can be optimized on-line by changing the number of basis functions and, in case of the complex filters, by changing the number of constituent primitive filters.

The hierarchical filter scheme is not tied to steerable filters of course. However, the use of steerable filters for the first step has the advantage that all orientations (and other deformations) of the primitive kernels are available, making the synthesis of the complex kernels more flexible and more exact. If only one type of complex kernel would be of interest and convolved with the whole image there is no or little advantage in the hierarchical filtering scheme. However, it is powerful if many different complex kernels are involved because all of them are synthesized from the same relatively small set of basis functions.

One application of the filter scheme that we demonstrated is the verification of the detected keypoint candidates. In earlier work this task has been performed by a neural network classification approach [4]. In contrast to the neural network approach the hierarchical filter approach has the benefit that explicit knowledge and models can be applied and tested. The filter scheme allows to extract with reasonable costs the low level features that are necessary to compare the image structures to the model.

REFERENCES

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

[2] R. Herpers, M. Michaelis, L. Witta, and G. Sommer, *Detection of keypoints in face images*, Tec.-Rep. GSF-Bericht 23/95, GSF-Forschungszentrum, Oberschleissheim, Germany, 1995

[3] R. Herpers, M. Michaelis, L. Witta, and G. Sommer, *Context based detection of keypoints and features in eye regions*, Proc. ICPR'96 Vol. II, Wien, Österreich, IEEE Computer Society Press, 23-28, 1996.

[4] R. Herpers, L. Witta, J. Bruske, and G. Sommer, *Dynamic cell structures for the evaluation of keypoints in facial images*, appears in Int. J. of Neural Systems, Spring 1997.

[5] Kamel et al., *Face recognition using perspective invariant features*, Pattern Rec. Letters **15**, 877–883, 1994.

[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, D-24105 Kiel, Germany, 1995.

[8] P. Perona, *Deformable kernels for early vision*, IEEE PAMI **17**, 488-499, 1995.

[9] A.L. Yuille, D.S. Cohen, and P.W. Hallinan, *Feature extraction from faces using deformable templates*, Proc. of IEEE Conf. CVPR '89, 104-109, 1989.