GAZE: An Attentive Processing Strategy to Detect and Analyze the Prominent Facial Regions*

R. Herpers[1], H. Kattner[1], H. Rodax[1], and G. Sommer[2]

[1] GSF-Institut für Medizinische Informatik und Systemforschung, Medis
D-85764 Oberschleißheim, Germany
E-mail: herpers@gsf.de

[2] Institut für Informatik, Lehrstuhl für Kognitive Systeme
Christian-Albrechts-Universität Kiel, D-24105 Kiel, Germany

Abstract

A new approach to a dynamic analysis of facial images is presented. Motivated by the eye movement strategies of the human visual system a computer based attentional mechanism is developed to recognize the prominent facial regions in natural human face images. The attentional mechanism is based on a feature representation, which is derived from gray-level face images by applying appropriate filter techniques. In a first processing step the most salient facial regions like the eyes, the nose, the mouth and the ears are sequentially localized. Subsequently, the detected and now spatially bounded regions are analyzed with more expensive methods. The exact positions of some relevant keypoints of these regions are determinable and a local interpretation of the detected image areas is derivable. By evaluating the different locally derived classification results a global interpretation of the whole face image can be generated. The presented attentive processing strategy is noted by a high degree of invariance properties. The detection of the prominent facial regions is independent of the exact position and the orientation of the face and of the facial components within the image. A high degree of scale invariance is achieved using a multi-scale representation of the saliency map for the search of the attentive regions. In addition, the proposed search algorithm is able to handle variable preconditions related to the illumination and the brightness contrast of the used face images.

1 Introduction

The visual search task in real world images is in the bottom-up case a NP-complete problem dependent on the image size [11]. A task-directed search based on selective attentional mechanisms can be computed in linear time complexity dependent on the number of objects to be searched in the image [11]. These complexity considerations suggest that attentional mechanisms are necessary to successfully solve an image analysis problem in real time. The application of attentional mechanisms generates a sequential order of spatially limited image regions and offers additional information about the considered objects [8].

Motivated by the eye movement strategies of the human visual system [12] an attentional mechanism selects only prominent regions in images [4]. Using attentional strategies in image processing systems only this information needs to be analyzed which is necessary for the given search task. Irrelevant image regions are ignored and do not bound additional computing resources. Thus, image processing with attention control simplifies computation and reduces the amount of processing [3, 7].

Elementary visual features like motion, color, orientation, edge information and others are derived in the human visual system at the early stage of the preattentive processing [10]. These features establish a preattentive representation upon which the control of the visual attention is based on. So, the regions are foveated in the order of their importance.

This paper presents a technical realization of an attentional mechanism localizing and analyzing the prominent facial regions in high resolution gray-level face images. The fundamental idea of the presented approach is an iterative strategy which is able to foveate dynamically the salient regions in the order of their importance. In comparison with classical recognition systems, the presented approach is based on a different recognition strategy which reduces the recognition problem to the analysis of only spatially limited but most important regions. It combines a cyclical as well as a hiearchical proceeding of the recognition process by applying attention strategies. It is shown, that the interpretation of a whole face image can be reduced to the analysis of several, spatially limited image regions, which are marked, however, by a high degree of saliency.

The iterative, attentive localization strategy also encloses the possibility of a relocalization of salient regions for several times. Beginning with the detection and analysis of local image areas, the information about the considered scene are improved iteratively. So, the amount of information being required to solve the recognition task is collected stepwise by each analyzed image part. The detailed information about each detected region increases and gets evaluated from processing step to processing step.

The methods which are applied during the several processing steps are mainly dependent on the actually available information. The global interpretation of the whole image should be computed at the very end of the processing by using all the derivable local information. Therefore, the presented computer based attentional mechanism is realized by different processing steps (fig. 1a).

In the first processing step spatially limited image regions are detected by using only simple image information, like edge information. Only simple features of the located regions are interpreted to decide about the necessity of applying expensive analysis methods [5]. In the second step the now spatially limited regions, in the following called foveated regions, are analyzed by advanced methods. Special filter tech-

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

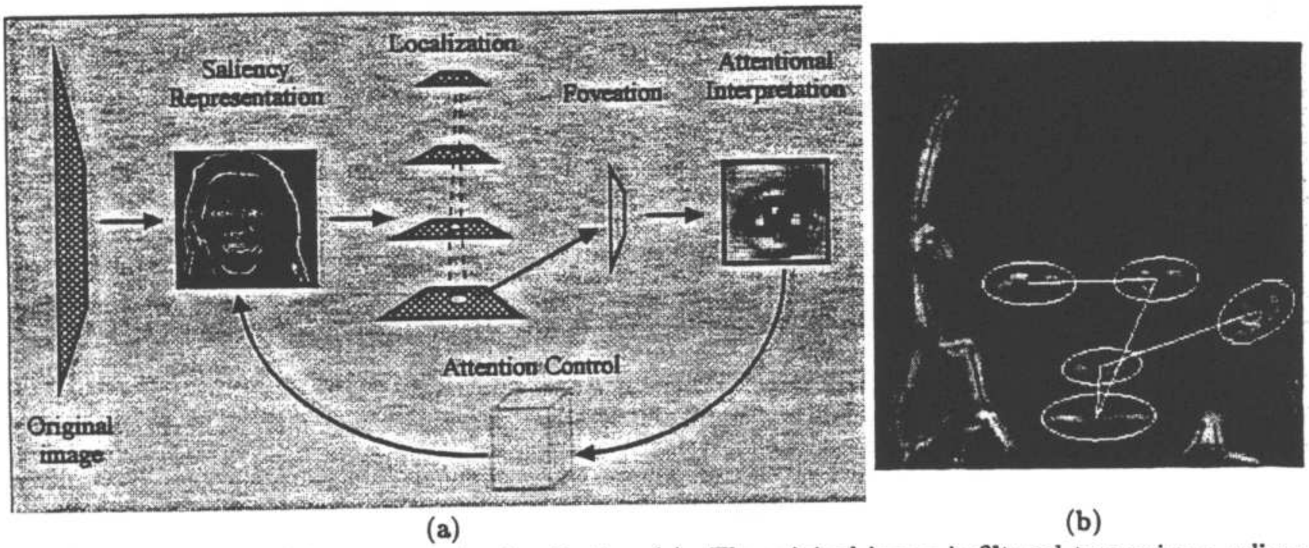


Figure 1: Proceeding of the preattentive localization (a). The original image is filtered to receive a saliency representation (b), which is scaled in different resolution levels. Beginning at the lowest resolution the most salient area is detected and projected to the next higher resolution level. The foveated region is analyzed applying detailed processing methods during an attentive processing step. The localization process is controled by an integrated attention strategy. Applying this proceeding sequentially to all salient image regions a scan path comparable to an eye movement record occurs (b).

niques are applied to detect characteristic keypoints in the foveated image regions, e. g. the center of the iris and the eye corners in an eye area. These keypoints are especially considered in a third step of the processing, applying steerable filters to analyze and classify them [6]. The information derived from these characteristic keypoints is evaluated to verify or to change the previous classification results. All the collected information of the already detected and interpreted regions and keypoints is used to control the further application of the attentional mechanism. This attention control implies a spatial estimation of different regions, in the following called anticipation, at predicted positions dependent on the already located regions.

Inspired by research of J.K. Tsotsos and S.M. Culhane, who developed 'a prototype for a data-driven visual attention' [4], the presented approach also demonstrates a data-driven attentional mechanism which is supported by model knowledge received during the runtime. Certain aspects of their model are not addressed in this paper, such as the inhibition of not relevant regions or the neural like modeling. Instead, emphasis is placed on a fast data-driven realization of a strategy localizing and analyzing prominent facial regions. Our attentional mechanism will be used as an essential component in building an image processing system classifying faces of children with dysmorphic abnormalities [9].

In the following chapter the theory and the developed methods of the localization mechanism are presented. Subsequently, the developed filter techniques to detect and verify the region dependent keypoints are discussed at an eye region. The success of the invariant localization of the prominent facial regions is shown at different examples in the last chapter.

2 Preattentive localization

Biological attentional mechanisms are based on various feature representations carrying all the information which is needed to control the visual attention or to generate a reaction of the visual system [8]. In the case of detecting prominent facial regions in gray-level images, the edge information is sufficient to guide the computer based attention [1, 5]. Therefore, filter responses of a first and second derivative of Gaussian in three orientations are calculated. These filters denoted by f_l are only applied on one particular scale caused by effiency and by reduced computational costs. Subsequently the filter responses are coupled with a 'control representation' C and scaled in a multi-scale representation [2]. This simple proceeding is acceptable because only qualitative properties of the filter responses are important for a feature based localization process [1]. Nevertheless this proceeding also supports an efficient realization of the presented computer based attention mechanism because first the features can be derived, then a 2-dimensional attention control representation can be calculated and combined with the feature map and, at last, the resulting saliency map is represented at different scales.

2.1 Saliency representation

The realization of the attentive localization mechanism is based on a 2-dimensional 'saliency representation' $S(t)$ in which the spatial distribution of the actual saliency of the underlying image is encoded. This saliency representation $S(t)$ is generated from the 'feature representation' U and the time dependent 'attention control representation' $C(t)$.

$$S(t) := U \times C(t) \tag{1}$$

The time t is used here as a discrete time step with $t := 0, .., \tau$. It is related to exact one localization step. The next time step $t+1$ begins when the localization of a new attentive region is started.

The feature representation U is defined as the weighted sum of the filter responses of the applied filters f_l to the image $I(X)$:

$$U := \sum_l a_l \ (f_l \otimes I(X)) \tag{2}$$

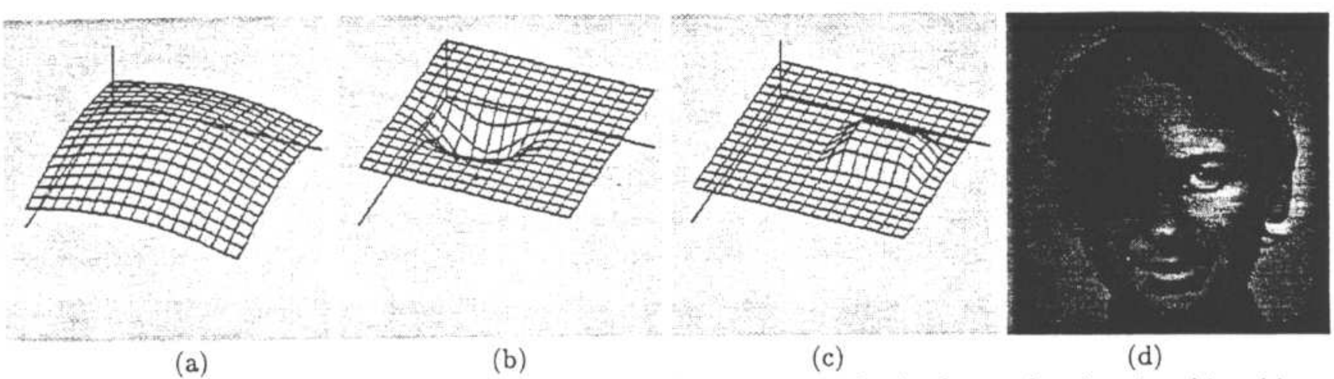


Figure 2: Weight maps used for the attentive control. A constant emphasis of central regions is achieved by a static weight map P (a). The suspension of the already foveated regions is modeled by suspension map $SR(t)$ (b). An expectation is generated by an antizipation map $A(t)$ (c). All pixelwise calculated weights are based on 2-dimensional Gaussian functions to achieve smooth transitions from increased to decreased areas. The resulting 2-dimensional control representation $C(t)$ build of all three components is linked to the original image (d). It is demonstrated that the already detected left eye has visibly a reduced saliency while the saliency of the right eye is increased.

when a_l are the corresponding weights and $I(X)$ the image with pixel positions $X := (x, y)^T$.

All these representations are weight matrices, in the following also called 'maps'. The particular saliency $s(X, t)$ at a position (X) is calculated by pixelwise multiplication of the feature value $u(X)$ and the control value $c(X, t)$.

$$s(X,t) := \; u(X) \; c(X,t) \qquad (3)$$

Attention control representation

The attention control representation $C(t)$ controls the attention process of the localization algorithm.

$$C(t) := P \times SR(t) \qquad (4)$$

The control representation $C(t)$ includes an emphasis of a priori known salient regions using a constant 'a priori weight map' P. Using this weight map P, predefined salient regions in an image can be emphasized. In the presented application investigating frontal face images the saliency of central image areas is increased against areas at the image corners, which have predefinedly a less saliency (fig. 2a).

Suspension representation

The second component of the saliency map $S(t)$ is a suspension map $SR(t)$ generated to suspend already foveated and analyzed regions from further processing steps. After the foveation of a region the saliency or attentive stimuli of this image area are reduced for the following localization steps (fig. 2b). This suspension of already localized regions is defined as a temporary local reduction of the activity of the salient features only related to the foveated region and its environment (fig. 2c).

2.2 Localization algorithm

The generated saliency representation is the basis of all following localization processes. The selection of the salient regions starts at the coarsest scale k of the saliency representation $S_k(t)$ employing a maximum search algorithm (fig. 3a). The maximum element $s_{k_{max}}(X, t)$ of the coarsest saliency map $S_k(t)$ is defined as:

$$s_{k_{max}}(X,t) := \max_{Y}\{s_k(Y,t)\} \qquad (5)$$

at the initial time step $t = 0$ with the pixel position $X := (x, y)^T$ and index k stands for the coarsest scale of the saliency map.

Detection of elliptical regions

In the proposed localization algorithm elliptical regions are detected by fitting a 2-dimensional Gaussian function to the distribution of the local saliency intensity (fig. 3b). A certain contour line h is taken to determine the boundary of an elliptical region.

$$(X - \bar{m})^T Cov^{-1}(X - \bar{m}) = h = const \qquad (6)$$

where $\bar{m}$ is the expectancy of the 2-dimensional Gaussian function or the center of the ellipse, Cov is the covariance matrix of all map elements or pixels of the localized region and h the contour line. The parameters of the fitted ellipse are updated iteratively to optimally cover the underlying feature representation at each resolution level (fig. 3c).

Region adaption

The adaption of the foveated region is computed with respect to every resolution level i of the processing hierarchy. The extent of the resulting ellipse, i.e. the length of the semi-axes of the ellipse a and b, is independent of the maximum intensity of feature representation. It only depends on the variances of the principal components of a spatially restricted saliency representation encoded in the covariance matrix $Cov_{i,j}$ and in the expectancy $\bar{m}_{i,j}$.

$$\bar{m}_{i,j} := \frac{1}{I} \sum_{X \in DilE_{i,j-1}} s_i(X) X \qquad (7)$$

$$Cov_{i,j} := \frac{1}{I} \sum_{X \in DilE_{i,j-1}} s_i(X) X X^T - \bar{m}_{i,j} \bar{m}_{i,j}^T \qquad (8)$$

$$\text{with} \qquad I := \sum_{X \in DilE_{i,j-1}} s_i(X) \qquad (9)$$

where $DilE_{i,j-1}$ is an expanded set of pixels lying in the elliptical region (defined below), the index i denotes the resolution level and the index j denotes the iteration set regarded to the resolution level i.

The extent of preliminary detected regions is expanded or reduced in a way that the size of the region

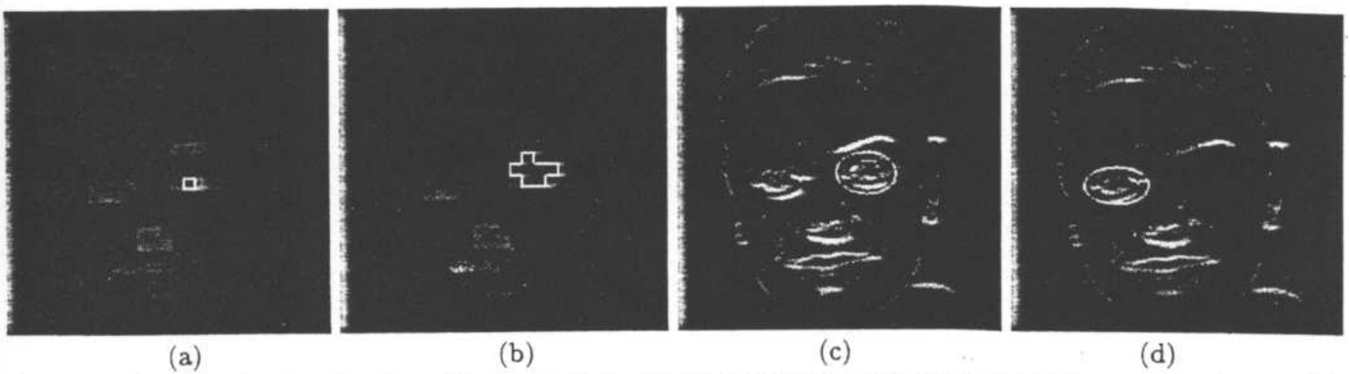


Figure 3: Preattentive localization of two prominent facial regions by evaluating the saliency maps in a multiscale representation. Maximum search of that element of the minimal saliency map with the highest saliency (a). The maximum element is expanded to the extension of the feature representation of the underlying eye region at the coarsest scale (b). Final selection of the left eye at the highest resolution level (c). After an detailed analysis, the foveated eye region is suspended from the following localization steps (d). The local analysis enables an anticipation of the second eye which is intended to be detected (d). The detected regions are spatially well adapted to the scale, extend and orientation of the eye.

i agrees with the extent of the feature representation of the associated object (fig. 3c). After an initial expansion of the maximum element $s_{k_{max}}(X,t)$ to a 3×3 neighborhood the detected maximum region is expanded to the underlying feature extension applying a two step elliptical region-growing algorithm.

Elliptical region-growing algorithm

In the first step, a set $E_{i,j}$ of pixels lying in the elliptical region is calculated:

$$E_{i,j} := \{X \in DilE_{i,j} | \quad (X - \bar{m}_{i,j})^T Cov_{i,j}^{-1} (X - \bar{m}_{i,j}) \leq h_i\} \tag{10}$$

where $\bar{m}_{i,j}$ is the center of the ellipse and $Cov_{i,j}$ is the covariance matrix regarded to the considered set. The calculation of this set is equivalent to the calculation of a fixed Mahalanobis-distance h_i from the center $\bar{m}_{i,j}$ and it implies an adaption of the calculated ellipse to the underlying intensity distribution.

In the second step, this set is expanded by extending the semi-axes of the calculated ellipse in steps of pixel units. Transforming the pixelwise extend of the semi-axes to the eigenvalues $\lambda_{1,2}$ of the covariance matrix $Cov_{i,j}$ a new covariance matrix $DilCov_{i,j}$ called 'dilated covariance matrix' is achieved. Calculating the corresponding dilated set $DilE_{i,j}$ to the dilated covariance matrix $DilCov_{i,j}$ a new base set for the next iterative adaption step is generated

$$DilE_{i,j} := \{X \in S_i(\tau) | \quad (X - \bar{m}_{i,j})^T DilCov_{i,j}^{-1} (X - \bar{m}_{i,j}) \leq h_i\} \tag{11}$$

where $S_i(\tau)$ is the definition set for all pixel positions of the resolution level i of the saliency representation at the particular time step τ. Subsequently, the new expectancy $\bar{m}_{i,j}$ and the actual covariance matrix $Cov_{i,j}$ is calculated refering to the dilated set. This region adaption is continued until the region is well adapted to the underlying saliency representation. This is reached when the breakcondition is fulfilled, or when, in other words, the center of the located ellipse $\bar{m}_{i,j}$ and the corresponding covariance matrix $Cov_{i,j}$ have not changed since the previous adaption step. Breakcondition:

$$(\bar{m}_{i,j} = \bar{m}_{i,j-1}) \wedge (Cov_{i,j} = Cov_{i,j-1}) \tag{12}$$

The coarse localized and restricted region is now projected to the next higher resolution level of the saliency representation. The iterative region adaption algorithm is computed separately at each resolution level to optimize the extent and the orientation of the considered region (fig. 3).

Suspension of regions

To compute the next attentive region, the already selected region has to be suspended from the following processing steps. A locally parameterized 2-dimensional Gaussian function, called 'suspension function' $SReg(t)$, is calculated for the detected region (fig. 3d). This proceeding produces smoothed transitions from already foveated regions to their surroundings. The suspension function is parameterized by the position and size parameters of the just located elliptical region R itself (fig. 2b).

$$\mathrm{SReg}(t) := 1 - e^{-\frac{1}{2}(X-\bar{m}_t)^T \mathrm{SCov}^{-1}(t)(X-\bar{m}_t)} \tag{13}$$

where $SCov(t)$ is the suspension covariance matrix which is equal to the last calculated covariance matrix $Cov_{i,j}$ of the adaption step of the previous localization process. $SCov(t) := Cov_{0,n_0}$ and $\bar{m}_t$ is the center vector of the just foveated region, where the highest resolution level is denoted by index $i = 0$ and the last iteration step of the adaption process is denoted by index $j = n_i$ (fig. 2b).

Suspension representation

The suspension representation $SR(t)$ or 'suspension map' is calculated out of all suspended regions of an image which have been foveated until time step t by evaluating the following recursive definition:

$$SR(t+1) := min(RF(SR(t)), \mathrm{SReg}(t)), \tag{14}$$

$$\text{and} \qquad SR(0) := \mathrm{SReg}(0) \tag{15}$$

with $RF(SR(t))$ is a refresh function which is responsible for a time variant suspension. After a fixed time delay D the saliency of the suspended region is gradually increased again. Therefore, the different salient regions may be relocalized for several times and results in a iterative strategy. Using a suspension functionality a sequential order of salient image regions occurs dependent on their salient importance encoded in the feature representation.

Refresh function

The refresh function is defined recursively for two cases, a linear and an exponential one:

$$RF_{lin}(SR(t)) := \ min(SR(t) + c_1, 1), \qquad (16)$$

$$\text{and } RF_{exp}(SR(t)) := \ SR(t) + c_2(1 - SR(t)) \ (17)$$

when $c_{1,2}$ are two different constants. Iterating this suspension functionality all salient image regions can be found (fig. 5).

3 Improved localization

After an initial 'orientation period', model knowledge is added to the attention mechanism to improve the localization of the prominent regions. In the proposed application, a face model is developed which contains the anthropometrical relations of the main facial regions like the eye, nose, mouth, and ear region. It is only used to guide the computer based attention to necessary and expected facial regions. In comparison to other related work, which uses the degree of accordance of a model to interpret a considered image part [1], the knowledge is only used here to support the localization process.

After a first initial phase of the localization algorithm one or more regions are detected and might be analyzed especially. If these detected regions are interpretable for their own, the face model of the knowledge base is adapted by these derived parameters to the actually given facial relations. The developed model is actualized after each successfull localization step with the derived parameters, like the scale, the orientation and the position of the main facial regions. So, a stepwise adaption of the model knowledge to the actually given individual structures is achieved and, therefore, the expectancy derived from the model is improved by each processing step.

Application of model knowledge to faces

The influence of the model knowledge to the localization process is realized in the presented application as following: If an eye region, for example, is detected in the first localization step without any knowledge support, and if this region is also interpretable as an eye region by applying detailed and only local analysis steps (see chapter 4), several parameters are possible to derive. The scale and orientation information as well as the position and extension of the detected face regions can be determined locally. These parameters are used to initialize the face model of the knowledge base. Evaluating this actual available facial information, the localization of other prominent facial regions can be supported by a derived expectation approach.

Anticipation representation

The knowledge influence to the attention control is realized similar to the suspension functionality of the attention mechanism. A 2-dimensional weight map is introduced, a so called 'anticipation map' $A(t)$. This representation is time dependent because it is calculated before each localization process. Applying this map, the saliency of the regions intended to be localized is increased gradually. The saliency of the rest of the image is decreased. Therefore, a 2-dimensional Gaussian distribution is calculated for each expected region in the same way as during a suspension process but with different sign. The Gaussian function is parameterized by the center of the expected region and by an estimated extension (fig. 2c). The anticipation map is the third component of the attention control representation $C(t)$ which is added after the initial orientation period. Therefore, formula 4 has to be replaced by:

$$C(t+1) := P \times SR(t) \times A(t) \qquad (18)$$

when $A(t)$ is the anticipation map. The anticipation map is updated after each localization step dependent on the results of a successful analysis step.

Attentive strategy

The underlying idea of the anticipation facility is to use the actual available knowledge about the image scene. The integrated model knowledge is only used to support the localization mechanism. The knowledge integration is thought as a top down strategy to control the computer based attention. This functionality reduces the unsharpness of the search process or, in general, the uncertainty of the recognition process. Applying the proposed strategy, the areas in which the expected image parts are to be searched are bound. It focuses the calculated attention only on usefull regions.

Proceeding of the attentive localization

The formal proceeding of the proposed attentive localization algorithm is now described in 12 different processing items. The attentive proceeding is relative independent of the derivation of the used features and it is robust against distinct derivations of the edge information.

1. **Derivation** of the feature representation U and initialization of the control map $C(0)$ with the a priori map P (fig. 2a). In the very first processing step, this map is identical with the initial attention control representation $C(0)$ by reason of the absence of a suspension. Subsequently the initial saliency map $S(0)$ is determined by combining the feature map U and the control map $C(0)$ (fig. 1b).
2. **Multi scale representation** of the saliency map $S_i(t)$ for $i := 0, .., k$ where the index $i = 0$ stands for the highest resolution and $i = k$ stands for the lowest. The localization algorithm starts at time step $t = 0$ cause of the initialization of the model knowledge. For all later localization processes $t := 1.., \tau$.
3. **Determination** of the maximum salient element $s_{k_{max}}(X, t)$ of the saliency representation at the coarsest resolution level $S_k(t)$ and selection of the 3×3 neighborhood (fig. 3a).
4. **Iterative expansion** of the detected region to the extend of the underlying saliency representation (elliptical region growing). Determination of the set $E_{i,j}$ and $DilE_{i,j}$. for $j = 1, .., n_i$ alternately until the extend and position of the located region doesn't change (fig. 3).
5. **Projection** of the detected region to the next higher resolution level $S_{i-1}(t)$ by scaling the parameter of the ellipse $Cov_{i,j}^{-1}$ (fig. 3b,c).

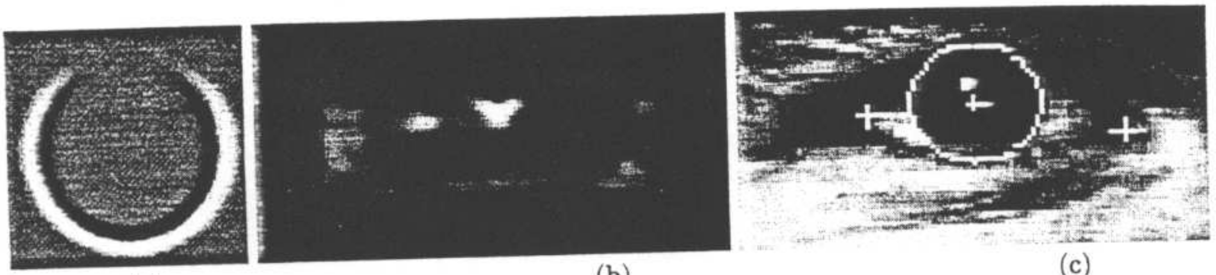


Figure 4: Detailed analysis of an eye region. A circular edge detector to localize the center of the iris (a). The upper part of the filter is missing because usually the upper boundary of the iris is not visible. The maximum of the scaled set of filter responses determines exactly the center of the iris (b). The corresponding scale of this optimal filter is used to calculate the radius of the iris (c). Subsequently the derived informations regarding scale and position information are used to support the determination of the position of the eye corners.

6. **goto** 4 until each resolution level has been considered. Descent of the resolution index $i := i-1$ for $i := k, .., 0$.
7. **Extraction** of the localized and now spatially well limited region (fig. 3b,c). Interface to detailed local analysis algorithms to derive region dependent parameters from the foveated region. Derivation of scale, orientation and semantic information from the foveated region (fig. 4).
8. **Initialization/update** of the face model. After the initial time step $t = 0$, the knowledge base is updated after each local analysis period.
9. **Derivation** of one or more expected regions from the actual state of the model knowledge. The generation of the anticipation map $A(t)$ is only executed if the face model is already initialized (fig. 2c).
10. **Suspension** of the just detected region by generating/updating the suspension map $SR(t)$ (fig. 2b and 3c,d).
11. **Calculation** of the new control representation $C(t+1)$ from the static, a priori map P, the new anticipation $A(t)$, and the actualized suspension map $SR(t)$ (fig. 2). Change from time step t to time step $t := t+1$. Link of the feature representation U with the new control representation $C(t+1)$ and generation of the saliency map $S(t+1)$ for the next localization process.
12. **goto** 2 until a complete interpretation of the image is received or a particular number of localization steps are executed.

4 Local analysis

The selected image regions are now especially analyzed dependent on a first local interpretation. Different relevant keypoints, which are mostly characteristic for a facial region, are searched by applying appropriate filter methods [6]. In the case of an eye region, for example, the existence and, therefore, the position of the iris and the eye-corners are very important (fig. 4c). In the presented example, a cicular filter is applied only to the foveated image part (fig. 4a). The filter responses determine exactly the center of the iris (fig. 4b). After the detection of the center, the boundary of the iris is segmented (fig. 4c). Now, the localization of the corresponding eye-corners are well determined because the spatial relations of the keypoint positions in the eye region are known. The morphological relations or morphometrical distances are known from investigations of many individual faces. The computation of the exact positions of the eye-corners is also carried out with the introduced attention strategy. Only predefined and spatially limited areas are investigated with more expensive and time consuming filter methods applying the already used steerability scheme [6].

Derivation of local parameters

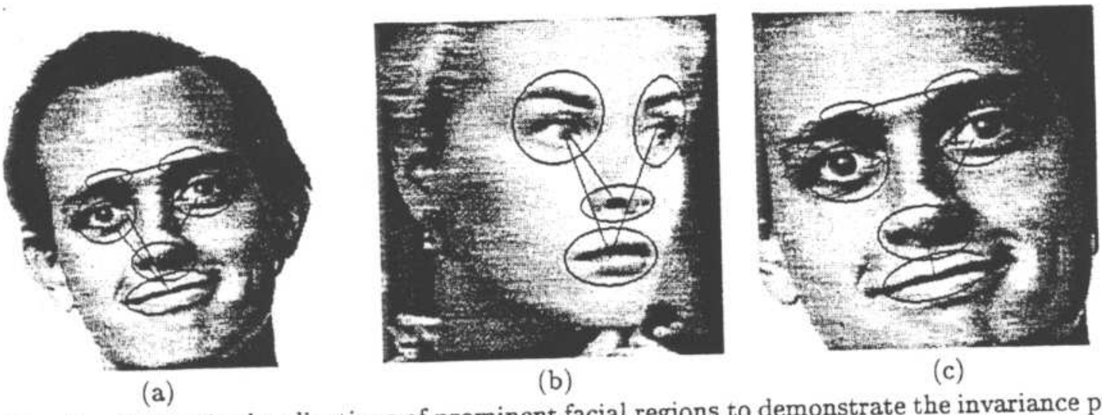
The application of special and time consuming filter methods is acceptable because the image areas are small and a first coarse idea or interpretation is known. The filters only cover a small range of scales given by the estimated scale of the considered image region. The derived parameters for the scale, orientation, and spatial position are used to initialize or to update the model knowledge. In addition, it is used further to improve detailed analysis steps. For example, after the recognition of the left eye, the scale and orientation parameter are used to improve the detection of the iris of the right eye.

5 Results

The presented methods are evaluated at more than 50 frontal face images with slightly different illumination conditions and camera positions (fig. 5). Some faces are tilt or slightly rotated in one direction. The scales and the brightness contrast of the recorded faces can also differ from image to image. At more than 95% of the face images both eye regions are detected and well restricted within the first 5 foveated regions. Including the anticipation facilities and using more expensive filter methods the percentage increases above 98% considering only the first 3 foveated regions. The reliable detection of eye region is relatively independent of their scale, their orientation and the brightness contrast (fig. 5).

The other facial regions like the nose, mouth, and ear region are mostly detected in later foveation steps. The spatial restrictions of these regions are also well executed with the exception of the ear region. The spatial limitation of the ear area is not reliable because of the high variability of the derived feature representation. However, the localization of the ear regions can be reliably guided by integrating the anticipation facility.

The application of the presented attentional strategy demonstrates a high degree of invariance properties concerning on the localization process and also on the recognition capability. The detection and classification of the prominent facial regions can be calculated independently of their position, their scale,



(a) (b) (c)

Figure 5: Results of attentive localizations of prominent facial regions to demonstrate the invariance properties. The localization of the several regions is invariant with respect to the tilt of the face (a). In addition, the localization is also independent of rotations in the X,Z-plane (b). The scale invariance property is demonstrated comparing figure (a) and figure (c) which have slightly different detected regions but a different scan path.

and their local orientation. As shown in the examples (fig. 5) the attentional mechanism achieves good rotation invariance properties concerning rotations in the X,Y-plane (fig. 5a) and rotations in the X,Z-plane (fig. 5b). A high degree of scale invariance performance could be integrated by computing the attentive regions using a multi-scale representation of the feature map (fig. 5c). Regions of different scales can be detected without changing the parameter configuration of the algorithms. In addition, the proposed localization algorithm is able to handle relatively variable preconditions related to the scale, illumination and brightness of frontal face images. There are no other preconditions to be taken like normalization or segmentation of the image data before starting the computation.

It is demonstrated that an artificial attention mechanism based on the calculation of edge features is able to detect the salient regions in face images. It is shown, that the interpretation of a whole face image can be reduced to the analysis of several, spatially limited image regions which are marked, however, by a high degree of saliency. The presented approach only uses the edge information as saliency input to find all attentive facial regions. These basic features are sufficient to detect and analyze the prominent facial areas in gray-level images, although in the attention system the calculation of other features is possible. The presented attentional mechanism is based on feature representations which are generated by a common set of basis functions, which are derived by a steerability scheme [6]. For the localization of the salient facial regions only a small subset of basis functions is needed. This first coarse representation is improved subsequently by adding more basis functions calculated only for the foveated region. The selected regions as well as the computed scan path in the investigated face images are in a good correspondence with the salient image areas. The proposed attentional strategy is also successfully tested on other real world images. However, the computed features have to be chosen appropriately.

Acknowledgments

We appreciate the support of Prof. Dr. S. Stengel-Rutkowski, of the Kinderzentrum Munich and Institute for Social Paediatrics and Youth Medicin providing the face images of the dysmorphic children and explaining us extensively the medical context of the considered dysmorphic abnormalities. The presented images in this contribution are mostly image data of volunteers of our institute because most of the dysmorphic face images are not allowed to be published. We are also grateful to Lars Witta, who helped to develop and to realize several components of the attentional mechanism. The work is supported by the DFG grants So 320\1-1 and Ei 322\1-1.

References

[1] R. Brunelli and T. Poggio, *Face recognition: features versus templates*, IEEE PAMI-15, no. 10, pp. 1042-1052, 1993.

[2] P.J. Burt, *Multiresolution techniques for image representation, analysis and smart transmission*, SPIE Conf. 1199, pp. 1-14, 1989.

[3] A. Califano, R. Kjeldsen and R. M. Bolle, *Data and model driven foveation*, Proc. 10th. ICPR 90, 1990.

[4] S.M. Culhane and J. K. Tsotsos, *A prototype for data-driven visual attention*, In: G. Sandini (ed.) ECCV'92, Springer, pp. 551-560, 1992.

[5] R. Herpers, H. Kattner, and G. Sommer, *GAZE: Eine attentive Verarbeitungsstrategie zum Erkennen und Lernen der visuell auffälligen Gesichtsregionen*, in: Mustererkennung 1994, W.G. Kropatsch et al. (Eds.), pp. 341-349, 1994.

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

[7] B. Ohlshausen, C. Anderson, and D. v. Essen, *A neural model of visual attention and invariant pattern recognition*, Tech. Rep CNS Memo 18, Caltech Pasadena, 1992.

[8] P.A. Sandon, *Simulating visual attention*, J. Cog. N. Sci, 2 (3),pp. 213-231, 1989.

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

[10] A. Treisman, *Preattentive processing in vision*, CVGIP Vol. 31, pp. 156-177, 1985.

[11] J.K. Tsotsos, *Analyzing vision at the complexity level*, Behavioral and Brain Sci., Vol. 13, pp. 423-469, 1990.

[12] A.L. Yarbus, *Eye Movements and Vision*, New York: Plenum Press, 1967.