# 1 Diagnosis Support of Patients with Facial Paresis

#### A. Gebhard, D. Paulus

Computers are used more and more in medical applications. Well-known examples are the analysis of radiographs or MR images [12, 9]. A special field is the analysis of human faces and facial features. Already realized applications exist e.g. for face recognition [1, 2] or face analysis [13, 6].

With the presented system we consider the problem of diagnosis support of patients with facial paresis. Approximately 350 patients per year are registered in the Department of Otorhino–Larygology of the University of Erlangen with new occurrences of this type of paresis (cf. [17] as an overview). The current way to diagnose the paresis is a subjective judgement of the functionality of the face muscles by a physician. The patient performs mimic exercises such as closing the eyes or showing the teeth, while he is observed by the physician. The subjective observations of the physician are then graded by means of two medical indexing systems [8, 14].

One part of the rehabilitation of the patients is to perform specific mimic exercises with the face's musculature. The success of the therapy is also observed by a physician or educated clinic personal. For every observation the patient has to travel to the clinics. Furthermore, educated personal is needed for the diagnosis and rehabilitation observation of the paresis.

The applications of our system are on the one hand diagnosis support in the clinics. We want to improve the subjective judgments of a physician by objective measurements and numerical features from the face. On the other hand the supervision of the rehabilitation process of the patient will be enhanced by placing the system to the patient's home. The patient can use the system in a convenient environment and does not have to travel to the clinics. The patient does *not* have to pay for the more convenient observation environment by wearing any artificial markers inside the face and he is allowed to move in front of the system's camera.

We present two different approaches for the analysis of facial features which will measure the face during the performance of mimic exercises. The result is a value for every picture with the level of asymmetry of the eye and mouth region which can be used to classify the facial paresis. The asymmetry of the face can be used as we handle just patients with paresis in one half of the face. This class of patients is the most frequent class. Patients with double sided pareses are approximately 1% of the face paresis patients.

This contribution is organized as follows: In Sect. 1.1 we compare the use of different data sources (2–D vs. 3–D images). A survey about related work on the field of facial image processing is given in Sect. 1.2. The localization of facial features will be described in detail in Sect. 1.3. It is followed by the description of the analysis approaches (Sect. 1.4). In Sect. 1.5 we show how the analysis results can be used to classify the facial paresis inside the observed regions. Results are shown in Sect. 1.6. Finally, Sect. 1.7 will recapitulate the contents of this paper with the main results in a short way.

## **1.1 Sensor Selection**

The analysis of facial images can be performed using either the frontal view in a 2–D projection, or taking into account the 3–D structure of the head. The first approach can be motivated by human perception. For human interaction and human interpretation of mimics, 2–D information seems to be sufficient. We can clearly tell whether a person smiles, frowns, or whether his eyes fixate ours, even when we close one eye. Paralyzed parts in the face are less perceived when they are in the lateral part of the cheeks. In the following we concentrate on 2–D images and use a model of the projected face.

Three-dimensional information on the face is available when stereo information is taken into account. However, it is disparity is difficult to estimate for facial images, since these images exhibit neither significant texture except for areas with facial hair — for area based matching, nor many lines or prominent points. Traditional stereo algorithms on the face thus result in sparse or coarse depth data.

## 1.2 Related Work

In this section we give a survey of related areas of the field of face image processing.

**Face Localization and Tracking** One of the basic components of many systems which process face images is a face localization module. Different approaches can be found in the literature. In the following we present exemplary two different approaches.

One approach was introduced by De Silva et al. [4]. A person is expected to sit in front of a homogenous background. A gray-level image of the head and the shoulders is captured. The edge strength representation of the image (cf. Sect. ??) is used to find an elliptic region containing a high amount of edge strength which is supposed to be the face.

Another approach was introduced by Oliver et al. [11] and bases on the segmentation of *blobs*. A blob is a "compact set of pixels that share a visual property that is not shared by the surrounding pixels" [11, p. 123], and is thus a special kind of region segmentation (Sect. ??). Oliver et al. use the normalized color as the visual property of a pixel. Every image coordinate is combined with color and brightness information by generating a vector  $(i, j, \frac{r}{r+g+b}, \frac{g}{r+g+b})^{\mathrm{T}}$ . A model for skin color was trained by thousands of samples that is valid for a broad spectrum of users. With an adaptive strategy, face color regions are grown on the complemented image data. The classification accuracy is close to 100%.

**Face Recognition** There are two major classes of face recognition operations. On the one hand there are methods using geometrical features or template-matching and on the other hand there is the processing on greylevel information.

The techniques basing on features or templates where analyzed in detail by Brunelli and Poggio [3]. They show the extraction of features or the template-matching with the goal of recognition. The results achieved by template-matching were better.

A prominent technique which bases on the processing of grey-level information is the *Eigenface* approach (e.g. [16]). The basic idea of the Eigenface approach is to encode a face image, and compare one face encoding with a database of models encoded similar. The encoding is done in the following way: A face image of size  $N \times N$  is interpreted as a point in an  $N^2$ -dimensional space. A set of training images of faces with similar overall configuration (e.g. face as the dominant image region) can be described by a relatively low dimensional subspace of the huge image space. Principal Component Analysis used with different images of one person gives vectors that best account for the distribution of face images within the entire image space. Those vectors are called *eigenfaces* (cp. the *eigenspaces* approach in Sect. ??).

For recognition, a new face image is transferred into its eigenface components and assigned to the eigenface with minimal distance. **Face Coding** The principal component representation of face images can also be used to encode faces, transmit the encoded information via a relative slow line, and decode and view them at the destination. Moghaddam and Pentland [10] propose this proceeding for the video telephony task.

Another model-based approach for this task was presented by Tao and Huang [15]. A basic articulation model can be influenced by articulation parameters such as facial action coding system (FACS) parameters of MPEG-4 facial animation parameters (FAPs).

**Medical Applications** In [7] a system for the automatic diagnosis of craniofacial dysmorphic signs is presented. Different multi-layer perceptrons were trained with a training set of 31 images using back-propagation as learning algorithm. The classification of a face whether craniofacial dysmorphic signs were present or not was done with a correct classification rate of 95%.

## 1.3 Localization and Tracking of Faces and Face Features

In the application presented in this contribution facial asymmetries are judged by means of asymmetries in the eyes, nose and mouth regions. The approaches we chose for the analysis of the facial paresis base on the localization of the mentioned facial features and of their surrounding (in case of the eyes the surrounding covers eyebrows and zygomatic). In our system the localization is done with a parametric face model, which is fitted to the image data by an energy-minimization process.

**Face Model** We assume that the patient's face is the dominant region inside the image. The background is expected to be either homogeneous or a background image is captured prior to the analysis which is used to part foreground from background. The localization is executed by calculating parameters of the face model shown in Figure 1. All calculations are performed on the edge strength part  $e_{\rm S}$  of a Sobel filtered image  $f({\rm Sect. ??})$ .

**Localization** The localization is performed as a four-step process. The *first* step is to localize the upper arc of the head. A ring segment with fixed width  $C_{\rm R}$  is found in the image which is expected to be the top of the patient's head. There are three parameters which have to be estimated: the



Figure 1: Parametric Face Model. All shown parameters are optimized during face localization.

x-coordinate  $x_{\rm h}$  and the y-coordinate  $y_{\rm h}$  of the origin  $q_{\rm h}$  and the radius  $R_{\rm h}$  of the arc. We expect much edge energy between patient and background. That lets us find the parameters of the circle and of the following part of the model in the edge strength representation of the original image. Equation 1 gives the edge-energy inside the head arc in the image:

$$E_{\rm h} = \int_{0}^{\pi} \int_{-R_{\rm h}}^{R_{\rm h}+C_{\rm R}} e_{\rm s}(x_{\rm h}+r\cos\phi, y_{\rm h}+r\sin\phi) \, dr \, d\phi. \tag{1}$$

To get the optimal model parameters the ratio of edge energy inside the head arc and the area of the head arc must be optimized. This optimization is done by the following maximization process:

$$\left(x_{h}^{*}, y_{h}^{*}, R_{h}^{*}\right) = \operatorname*{argmax}_{\left(x_{h}^{*}, y_{h}^{*}, R_{h}^{*}\right)} \frac{E_{h}}{\left(\frac{(R_{h}^{*} + C_{R}^{*})^{2} - R_{h}^{*}}{2}\right)^{C_{h}}}.$$
(2)

The constant  $C_{\rm h}$  influences the energy to area ratio. The values of the constants are  $C_{\rm R}=20$  and  $C_{\rm h}=1.4$ .

The optimization is performed by an adaptive random search with a subsequent local simplex method [5]. To speed up the optimization there are restrictions for plausible parameters: The parameter  $x_{\rm h}$  can vary form  $N_{\rm y}/4$ to  $3 N_{\rm y}/4$ ,  $y_{\rm h}$  form M/4 to M/2, and  $R_{\rm h}$  from  $N_{\rm y}/10$  to  $N_{\rm y}/3$ .

The next step is to localize the ears which are modeled as rectangles positioned below the arc of the head. The parameters which have to be determined are the position of the origin  $q_{ea}$ , the width  $W_{ea}$  and height  $H_{ea}$  of the ears, and the distance  $2D_{ea}$  between the two ears. The equation to calculate the edge-energy inside the ears' region is

$$E_{ea} = \int_{0}^{W_{ea}} \int_{0}^{H_{ea}} \{e_{s}(x_{ea} + w + D_{ea}, y_{ea} + h) + e_{s}(x_{ea} + w - D_{ea}, y_{ea} + h)\} dh dw$$
(3)

and similar to 2 the calculation of the optimal ear parameters bases on the ratio of edge energy inside the ears regions and the area of this regions. They can be determined as

$$(x_{ea}^{*}, y_{ea}^{*}, D_{ea}^{*}, W_{ea}^{*}, H_{ea}^{*}) = \underset{(x_{ea}, y_{ea}, D_{ea}, W_{ea}, H_{ea})}{\operatorname{argmax}} \frac{(E_{ea} D_{ea})^{C_{ea}}}{(W_{ea} H_{ea})} \quad (4)$$

with the ratio influencing constant  $C_{ea} = 1.4$ . Additionally the distance between the left and right ear region is involved in the optimization as they are used as the horizontal boundaries of the face and therefore the distance  $D_{ea}$  should be as big as possible.

There are anatomic restrictions for the ear parameters. The origin of the ears  $q_{\rm ea}$  must have a lower horizontal distance than  $R_{\rm h}/3$ , the vertical position of  $q_{\rm ea}$  must be below the origin of the head arc  $q_{\rm h}$ , but with a lower distance than  $R_{\rm h}/2$ .  $D_{\rm ea}$  must be between  $0.9R_{\rm h}$  and  $N_{\rm y}/3$ .

The eyes, which are found in the *third* step, are modeled as ellipses. The parameter  $\mathbf{q}_{ey} = (x_{ey}, y_{ey})^{T}$  is the position of the origin of the eyes,  $A_{ey}$  and  $B_{ey}$  the length of the ellipses axis,  $2H_{ey}$  the vertical and  $2D_{ey}$  the horizontal distance of the eye centers to each other. We use the following equation to calculate the edge energy inside the eyes' regions:

$$E_{\rm ey} = \int_{0}^{2\pi} \int_{0}^{1} e_{\rm s}(x_{\rm ey} + A_{\rm ey}r\cos\phi + D_{\rm ey}, y_{\rm ey} + B_{\rm ey}r\sin\phi + H_{\rm ey}) + e_{\rm s}(x_{\rm ey} + A_{\rm ey}r\cos\phi - D_{\rm ey}, y_{\rm ey} + B_{\rm ey}r\sin\phi - H_{\rm ey}) dr d\phi.$$
(5)

With the following optimization we get the eye parameters:

$$(x_{ey}^{*}, y_{ey}^{*}, D_{ey}^{*}, A_{ey}^{*}, B_{ey}^{*}, H_{ey}^{*}) = \underset{(x_{ey}, y_{ey}, D_{ey}, A_{ey}, B_{ey})}{\operatorname{argmax}} \frac{E_{ey}}{(A_{ey}^{2} + B_{ey}^{2})^{C_{ey}}}$$
(6)

with a constant  $C_{ey} = 1.4$  influencing the ratio of energy to area of the eyes' regions.

The anatomic restrictions here are: The horizontal distance of the origin of the eyes  $q_{\rm ey}$  must be less than  $R_{\rm h}/3$ , the vertical position must be greater than  $y_{\rm h}$  but lower than  $y_{\rm h} + R_{\rm h}$ . A<sub>ey</sub> and B<sub>ey</sub> must be lower than  $0.4R_{\rm h}$ and D<sub>ey</sub> lower than D<sub>ea</sub>. The two eye regions must not overlap and the eye regions must not overlap the ears' regions.

Finally the nose/mouth region is to be found. It is modeled as a triangle stump with parameters: origin  $\boldsymbol{q}_{\rm nm} = (x_{\rm nm}, y_{\rm nm})^T$ , height  $H_{\rm nm}$ , length of base line  $W_{\rm nm}$  and the length of top line  $T_{\rm nm}$ .

The amount of edge strength in the nose mouth region can be determined by the following equation:

$$E_{\rm nm} = \int_{0}^{H_{\rm nm}} \int_{-1}^{1} e_{\rm s}(x_{\rm nm} + w (W_{\rm nm} - \frac{h}{H_{\rm nm}} (W_{\rm nm} - T_{\rm nm})), y_{\rm nm} + H_{\rm nm}) \, dw \, dh$$
(7)

The optimal parameters are found as

$$(x_{\rm nm}^{*}, y_{\rm nm}^{*}, W_{\rm nm}^{*}, T_{\rm nm}^{*}, H_{\rm nm}^{*}) = \arg_{\substack{\text{argmax}\\(x_{\rm nm}, y_{\rm nm}, W_{\rm nm}, T_{\rm nm}, H_{\rm nm})}} \frac{(E_{\rm nm})^{C_{\rm nm}} H_{\rm nm}}{W_{\rm nm} + T_{\rm nm}}.$$
(8)

Here the constant to influence the ratio of edge energy to area is  $C_{\rm nm} = 1.2$ . The height of the nose/mouth region appears in the numerator of the ratio in (8) to avoid that the optimization result is a region that covers just the nostrils or the mouth and not both of them.

The horizontal distance of the origin  $q_{\rm nm}$  must be lower than  $D_{\rm ey}/3$  from  $q_{\rm ey}$ . The vertical position must be between  $y_{\rm ey} + 2D_{\rm ey}$  and  $y_{\rm ey} + 3D_{\rm ey}$ .  $W_{\rm nm}$  must be between  $2D_{\rm ey}$  and  $3D_{\rm ey}$ ,  $T_{\rm nm}$  between  $D_{\rm ey}$  and  $2D_{\rm ey}$ , and  $H_{\rm nm}$  between  $D_{\rm ey}$  and  $3D_{\rm ey}$ .

All the restrictions to the optimized parameters were imposed because of observations of anatomic facts. The constants  $C_{\rm R}$ ,  $C_{\rm h}$ ,  $C_{\rm ea}$ ,  $C_{\rm ey}$ , and  $C_{\rm nm}$  which were used for the calculations were determined experimentally by localization and tracking of patient faces and facial features in 1000 images. **Tracking** When the face and the facial features are localized in one image (i.e. the parameters of the face model are calculated) the face and features can be found in the following image in approximately the same position as we postprocess a video stream with 25 frames per second and a relatively slow moving patient.

We initialize the simplex optimization with the parameters from image i and a set of parameters which are normal distributed with mean old parameter and variance depending on the expected parameter variation.

That reduces the search area very much and we initialize the simplex optimization with less parameter sets (instead of 5000 we use 500) to get the optimal parameter in a reliable way. The reduction of the initialization set also results in a noticeable decrease of calculation time (instead of 40 sec we need 8 sec per image).

#### **1.4 Analysis of Facial Paresis**

As written in the introduction we operate with patients with single sided facial pareses. In this case asymmetries can occur in the eye and mouth region when specific mimic exercises are performed. This asymmetries are considered to be symptoms for the present paresis. We will give two different data driven analysis methods for the analysis of the eye region and mouth region of a human face.

The first mimic exercise is to lift up the eyebrows such that wrinkles will appear on the forehead. In the following we will call this exercise "frowning". Depending on the grade of paresis, some patients are not able to lift the eyebrow. This will result in certain asymmetries in the eye/eyebrow-region. Also the second exercise, the closing of the eyes, can generate asymmetries in this face area, as some patient are not able or have severe problems to close their eyes.

The other two exercises tell us something about the patient's mouth region. Asymmetries can arise when patients try to point their mouth or when they show the teeth. We record images of the patient while he is performing the mentioned exercises. To grade the asymmetries we need an additional view, the relaxed face.

**Comparison of Intensity Values and Edge Strength** The first attempt is the direct comparison of gray-level values of mouth and eyes. To get a feature of the eye region, we take the gray-values of the left eye, mirror the single lines and match the gray-levels with the right eye by varying the x- and y-coordinate to find the minimal absolute sum  $D_1^*$  of the pixel differences.

$$D_1^* = \min_{x_t, y_t} \int_{iight \, Eye} |(f(x, y) - f(2(B_{ey} - D_{ey}) - x + x_t, y - 2H_{ey} + y_t)| \, dx \, dy.$$
(9)

The absolute sum divided by the area of the right eye (in our face model both eyes have the same size) is taken as feature  $c_1$ .

$$c_1 = \frac{D_1^*}{(A_{\rm ey}B_{\rm ey})}.$$
 (10)

To analyze the mouth region we find the row index  $r_{min}$  inside the mouth region that will give the minimal sum of absolute differences when matching the pixels on the left of row  $r_{min}$  with those on the right side. Row  $r_{min}$ represents the vertical symmetric axis of the mouth which is the elongation of the nasal labial fold.

$$D_{2}^{*} = \min_{r} \int_{Mouth \ left}^{r} \int_{0}^{H_{\text{nm}}} |(f(x,y) - f(2r - x,y)| \ dx \ dy$$
(11)

 $D_2^*$  is devided by the area of the analyzed region to get a feature for the asymmetry of the mouth region.

$$c_2 = \frac{D_2^*}{r_{min} H_{\rm Nm}}.\tag{12}$$

To keep the following more predictable: Odd-indexed features  $c_{2i+1}$  result from comparisons of the eye regions, even-indexed features  $c_{2i}$  result from the mouth region. As mentioned before interesting areas inside the face are often regions where changes of the gray-values occur. Such regions appear emphasized in the edge-strength representation of the image. For that reason we additionally applied the methods not only on the gray-levels f but also on the edge-strength part  $e_{\rm S}$  of the Sobel-filtered image. That gives us the next two features  $c_3$  for the eye region and  $c_4$  for the mouth region. Using averaging filter responses for face analysis The second class of approaches for the analysis of facial asymmetries arises from the theory of steerable filters [18]. We use the response of averaging rotated wedge filters to characterize the direction information in the environment of certain key points (Figure 2). The key points here are the corners of the eyes and the mouth and the extracted information contains the opening angle of those facial features (Figure 3). The disadvantage of this approach is that the positions of the angles of the eyes and the mouth have to be determined as exact as possible. This additional localization is of course another source for errors. The localization of these features can be very hard even for a exercised person and it is often the case that just an unprecise estimation can be performed.

To get an estimate for the position of the angles of eyes and mouth we use the columns of the surrounding boxes of eyes and the mouth. We noticed that the angles often appear noticeable darker than the surrounding intensity values. The search for the angle positions is started at the columns of the outer borders of the localized facial feature regions (cf. Sect. 1.3). We calculate the average intensity value of one column, and compare the minimum of the column to it. If the quotient is below a threshold  $\theta_e$  of  $\theta_m$ , we stop the process and consider the angle as found. This simple method gives useful results which are used in the analysis process.

We take the localized eye angles and mouth angles as the key points of the wedge filters. The filters are rotated not the whole  $2\pi$  but only  $\pi$  over the face feature in 2 degree steps which gives 91 averaged gray-values for every localized four facial feature angle. The opening angle of the wedge is 6 degree.

The results are the four vectors  $\boldsymbol{w}_i, i = 1, \ldots, 4$  of filter responses with 91 entries each  $\boldsymbol{w}_i = [w_{i,1}, \ldots, w_{i,91}]^{\mathrm{T}}$ . Feature  $\boldsymbol{w}_1$  results from the filtering of the corner of the right eye,  $\boldsymbol{w}_1$  from the left eye,  $\boldsymbol{w}_3$  from the right corner of the mouth, and  $\boldsymbol{w}_4$  from the left corner. The filters are also applied to the edge-strength part  $e_{\mathrm{S}}$  of a Sobel-filtered image. That gives another 4 vectors  $\boldsymbol{w}_5$  to  $\boldsymbol{w}_8$  resulting from the respective facial feature corners. With these eight vectors we can calculate another eight features values to analyze the facial asymmetry.  $c_5$  is the absolute component difference sum of  $\boldsymbol{w}_1$  and  $\boldsymbol{w}_2$ .

$$c_5 = \sum_{i=1}^{91} |w_{1,i} - w_{2,i}| \tag{13}$$



Figure 2: a) an averaging mask centered at key point  $\boldsymbol{p}_k$ . b) the response of the individual wedge filters. c) the reconstructed (Gaussian smoothed) filter response

The same is done with  $\boldsymbol{w}_3$  and  $\boldsymbol{w}_4$  to calculate the feature  $c_6$ ,  $c_7$  with  $\boldsymbol{w}_5$ and  $\boldsymbol{w}_6$ , and  $c_8$  with  $\boldsymbol{w}_7$  and  $\boldsymbol{w}_8$ . The feature  $c_9$  is generated after matching  $\boldsymbol{w}_1$  to  $\boldsymbol{w}_2$ . This is done by translations  $(t_1)$  and scalings  $(s_1)$  of the vector components of  $\boldsymbol{w}_2$ :

$$c_9 = \min_{s_1, t_1} \sum_{i=1}^{91} |w_{1,i} - s_1 w_{2,i+t_1}|$$
(14)

Analog to the feature  $c_9$  the features  $c_{10}$ ,  $c_{11}$  and  $c_{12}$  can be calculated.

## **1.5** Classification of Facial Paresis

To classify a face whether a paresis is present or not, we proceed as follows. First we calculate the features  $c_1$  to  $c_{12}$  while the person's face is in the following states:

- 1. relaxed face: all face muscles in a relaxed state
- 2. lifting up the eyebrows: the result of this motion are wrinkles on the forehead



Figure 3: Smoothed responses of the wedge filters applied to the angles of the eyes and the mouth

- 3. closing the eyes: patients with paresis in the eye region have problems to do this
- 4. pointing of the mouth
- 5. showing the teeth: the last two exercises give information about a potential paresis in the mouth region

That gives five sets of the twelve parameters each which are used for classification:

- 1.  $c_{1,norm}, ..., c_{12,norm}$ : extracted parameters while the face is in a relaxed state
- 2.  $c_{1,frown}, ..., c_{12,frown}$ : extracted parameters while frowning
- 3.  $c_{1,close}, \ldots, c_{12,close}$ : extracted parameters while closing the eyes
- 4.  $c_{1,tip}, ..., c_{12,point}$ : extracted parameters while pointing the mouth
- 5.  $c_{1,teeth}, ..., c_{12,teeth}$ : extracted parameters while showing the teeth

To analyze the facial paresis we are interested in asymmetries which are caused by the performation of the mimic exercises. To subtract the asymmetry information from the extracted features which is caused not by the paresis but other factors like the illumination or anatomic reasons, all the parameters  $c_{1,frown}, ..., c_{12,frown}, c_{1,close}, ..., c_{12,close}, c_{1,tip}, ..., c_{12,tip}$ , and  $c_{1,teeth}, ..., c_{12,teeth}$  are normalized by the parameters  $c_{1,norm}, ..., c_{12,norm}$ , which mainly contain asymmetry information not generated by the mimic exercises.

This normalization is done by calculating the ratios

- 1.  $\frac{c_{1,frown}}{c_{1,norm}}, \dots, \frac{c_{12,frown}}{c_{12,norm}}$
- 2.  $\frac{c_{1,close}}{c_{1,norm}}, \dots, \frac{c_{12,close}}{c_{12,norm}}$
- 3.  $\frac{c_{1,tip}}{c_{1,norm}}, \dots, \frac{c_{12,tip}}{c_{12,norm}}$
- 4.  $\frac{c_{1,teeth}}{c_{1,norm}}, \dots, \frac{c_{12,teeth}}{c_{12,norm}}.$

The four sets of normalized parameters with 12 values each are finally used to detect a facial paresis. Again, the odd-indexed normalized parameters contain information about the eye region, even-indexed normalized parameters about the mouth region. In the present state of the system the normalized features are thresholded to get the information whether facial paresis exists or not. For every set of features we calculate if facial paresis is present. E.g. if  $\frac{c_{1,frown}}{c_{1,norm}}$ ,  $\frac{c_{1,close}}{c_{1,norm}}$ , or  $\frac{c_{1,teeth}}{c_{1,norm}}$  are greater than the thresholds  $\theta_{1,frown}$ ,  $\theta_{1,close}$ ,  $\theta_{1,frown}$ , or  $\theta_{1,close}$  we consider a paresis in the eye region. If  $\frac{c_{2,frown}}{c_{2,norm}}$ ,  $\frac{c_{2,coorm}}{c_{2,norm}}$ , or  $\frac{c_{2,teeth}}{c_{2,norm}}$  are over the thresholds  $\theta_{2,frown}$ ,  $\theta_{2,close}$ ,  $\theta_{2,point}$ , or  $\theta_{2,teeth}$  facial paresis in the mouth region is supposed.

This gives six ways to detect facial paresis in the eyes region and another six ways to diagnose the mouth region. The classification results of all twelve feature sets are presented in Sect. 1.6.

#### **1.6 Experimental Results**

The evaluation of the whole diagnosis support system was performed with 16 patients with different grades of paralysis and 4 healthy persons. In this section we present the results of all our experiments.

We started with the generation of image sequences of the 20 persons. The persons performed mimic exercises in front of a homogeneous background. and the image series including 20 images were taken when the extremal positions of the exercises described in the last section were reached. That gives five image sequences of length 20 of every person.

In every first image of the series the face was located by means of the parametric face model and tracked in the remaining 19 images. In Figure 4



Figure 4: Localization example of face and facial features

Head: $(x_{h}^{*}, y_{h}^{*}, R_{h}^{*})$	$(191,\!123,\!65)$
Ears: $(x_{ea}^*, y_{ea}^*, D_{ea}^*, W_{ea}^*, H_{ea}^*)$	$(192,\!144,\!20,\!62,\!93)$
Eyes: $(x_{ey}^*, y_{ey}^*, D_{ey}^*, A_{ey}^*, B_{ey}^*, H_{ey}^*)$	(194,153,31,26,22,1)
Nose/Mouth $(x_{nm}^*, y_{nm}^*, W_{nm}^*, T_{nm}^*, H_{nm}^*)$	(190, 229,  63,  31,  43)

Table 1: Computed parameters (see Figure 1) for real face

the localization results are shown graphically. In that example the input image was of size  $384 \times 288$ . The calculated parameters are shown in Table 1. Totally the localization was successful in 82% of the eyes and in 73% of the mouth and whenever the facial features where localized correctly the tracking was done error-free.

The localization of the eye and mouth angles (cf. Figure 5) was performed correctly with a rate of 50% for the eye angles and 45% for the mouth angles. In the case of an error the correct position was hand-segmented.

With the localized facial features the classification of the facial paresis



Figure 5: Localization example of the eye angle and mouth angle



Figure 6: Results of face analysis: a) Patient, left: feature  $c_1$  (eye), right: feature  $c_2$  (mouth); b) Healthy Person, left: feature  $c_1$  (eye), right: feature  $c_2$  (mouth)

was performed.

In Figure 6 characteristical evaluations of the features  $c_1$  (eye region) and  $c_2$  (mouth region) are shown. The features were calculated while the observed person performed the five mimic expressions described in the last section. The numbers at the *x*-axis show the exercise which was performed to calculate the feature. A patterned entry shows a change of asymmetry that might be originated by a facial paresis.

Then we generate the ratio of the parameters belonging to the relaxed face with the corresponding other ones. That produces for every of the twelve features  $c_1$  to  $c_{12}$  four ratios. The ratios are compared with a threshold and if a certain number of ratios are greater than the threshold the face is classified to contain paresis or not.

The extracted features were used to classify face images into healthy persons and patients with facialis paresis. The selected threshold for all ratios was  $\theta = 0.7$ . Table Table 2 shows the classification results. The correctness of a classification was ermittelt by comparation with the grading of a medical specialist.

Eye						
	$c_1$	$c_3$	$c_5$	$c_7$	$C_9$	$c_{11}$
healthy	0.75	0.75	0.75	0.50	0.75	0.25
$\mathbf{paresis}$	0.6	0.6	0.67	0.67	0.73	0.73
Mouth						
	$c_2$	$c_4$	$c_6$	$c_8$	$c_{10}$	$c_{12}$
healthy	1.0	1.0	0.25	0.25	0.50	0.50
$\mathbf{paresis}$	0.87	0.87	0.73	0.73	1.0	0.93

Table 2: Classification results

## 1.7 Conclusion

We presented the basics of a system for diagnosis support and rehabilitation supervision of patients with facial paresis. The advantages of this system are an objective method to diagnose facial paresis. This is performed by means of measurements and numerical features from the face.

In a first step we localized the face and facial features using a dynamic face model and a energy maximization Simplex approach. The segmented model parameters where used to analyze the eye and mouth regions of the face towards asymmetries as asymmetries are symptoms for facial paresis. The extracted numerical features made us grade the facial paresis in mouth and eye regions. We evaluated the system by 15 patients and 5 healthy persons.

- $e_{\rm S}$  Edge Image Strength page.4
- f Image function page.4
- $C_{\mathrm{R}}$  Ring Width page.4
- $x_{\rm h}$  Origin page.4
- $y_{\rm h}$  Origin page.4
- $\boldsymbol{q}_{\mathrm{h}}$  Origin head page.4
- $R_{\rm h}$  Radius page.4
- $E_{\rm h}$  Face energy term page.5
- $C_{\rm h}$  Constant for head page.5
- $N_{\rm Y}$  Image Size Y page.5
- M Image size Y page.5
- $q_{ea}$  Origin ears page.6
- $W_{ea}$  Width eyes page.6
- $H_{ea}$  Height ears page.6
- $D_{ea}$  Distance ear page.6
- $E_{ea}$  Ears energy term page.6

 $x_{ea}$  Origin page.6

- $y_{ea}$  Origin page.6
- $C_{ea}$  Constant for ears page.6
- $\boldsymbol{q}_{\mathrm{ev}}$  Origin eyes page.6
- xey Origin page.6
- yey Origin page.6
- $A_{\rm ey}$  Ellipse par 1 for eye page.6
- $B_{\rm ev}$  Ellipse par 2 for eye page.6
- $H_{\rm ev}$  Height eyes page.6
- $D_{\rm ev}$  Distance eyes page.6
- $E_{\rm ev}$  Eyes energy term page.6

- $C_{\rm ev}$  Constant for eyes page.6
- $q_{\rm nm}$  Origin nose mouth page.7
- $x_{\rm nm}$  Origin page.7
- $y_{\rm nm}$  Origin page.7
- $H_{\rm nm}$  Height of nose/mouth page.7
- $W_{\rm nm}$  Width of nose/mouth page.7
- $T_{\rm nm}$  Width of nose/mouth (top) page.7
- $E_{\rm nm}$  Nose/Mouth energy term page.7
- $C_{\rm nm}$  Constant for mouth/nose page.7
- $D_1^*$  Pixel difference page.8
- $D_1^*$  Pixel difference eyes page.9
- c Face Feature page.9
- $D_2^*$  Pixel difference mouth page.9
- $\boldsymbol{w}$  Wedge filter feature vector page.11
- w Wedge averaged value page.11

## References

- V. Bruce, P. Hancock, M. Burton: Comparison between Human and Computer Recognition of Faces, in ICAFGR, Nara, Japan, 1998, S. 408– 413.
- [2] R. Brunelli, T. Poggio: Face Recognition through Geometrical Features, in G. Sandini (Hrsg.): Computer Vision — ECCV '92, Lecture Notes in Computer Science, Springer, Santa Margherita Ligure, 1992, S. 792–800.
- [3] R. Brunelli, T. Poggio: Face Recognition: Features versus Templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, Bd. 15, Nr. 10, 1993, S. 1042–1052.
- [4] L. De Silva, K. Aizawa, M. Hatori: Detection and Tracking of Facial Features, in SPIE Visual Communications and Image Processing'95

(VCIP'95), Bd. 2501, The International Society for Optical Engineering, Mai 1995, S. 2501/1161-2501/1172.

- [5] S. M. Ermakov, A. A. Zhiglyavskij: On Random Search of Global Extremum, Probability Theory and Applications, Bd. 28, Nr. 1, 1983, S. 129–136.
- [6] R. Herpers, H. Kattner, H. Rodax, G. Sommer: GAZE: An Attentive Processing Strategy to Detect and Analyze the Prominent Facial Regions, in M. Bichsel (Hrsg.): Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition - IWAFGR'95, Zurich, 1995, S. 214-220.
- [7] R. Herpers, H. Rodax, G. Sommer: A neural network identifies faces with morphological syndrome, in S. Andreassen (Hrsg.): Artificial Intelligence in Medicine, IOS Press, Amsterdam, 1993, S. 481-485.
- [8] J. W. House: Facial Nerve Grading Systems, Laryngoscope, Bd. 93, 1983, S. 1056-1069.
- [9] G. Lohmann, D. von Cramon: Automatic Detection and Labelling of the Human Cortical Folds in Magnetic Resonance Data Sets, in ECCV98, Freiburg, 1998.
- [10] B. Moghaddam, A. Pentland: An Automatic System for Model-Based Coding of Faces, 317, MIT Media Lab Vismod, 1995.
- [11] N. Oliver, A. Pentland, F. Berard: LAFTER: Lips and Face Real Time Tracker, in IEEE Conference on Computer Vision and Pattern Recognition, S. Juan, Puerto Rico, 1997, S. 123-129.
- [12] M. Pelka, K.-H. Kunzelmann, D. Paulus, A. Winzen: Automatic Digital Subtraction Radiography using Simulated Annealing, in IADR, Seattle, 1994.
- [13] A. Samal, P. Iyengar: Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey, Pattern Recognition, Bd. 8, 1992, S. 65–77.
- [14] E. Stennert, C. H. Limberg, K. P. Frentrup: Parese- und Defektheilungs-Index, HNO, Bd. 25, 1977, S. 238-245.
- [15] H. Tao, T. Huang: Deriving Facial Articulation Models From Image Sequences, in Proc. IEEE Intern. Conf. on Image Processing, ICIP'98, Chicago, IEEE, Okt. 1998, S. III 158–161.
- [16] M. Turk, A. Pentland: Face Recognition Using Eigenfaces, in Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawai, Juni 1992, S. 586-590.
- [17] S. Wolf, M. M"uller, W. Schneider, C. Haid, M. Wigand: Facial Nerve

Function after Transtemporal Removal of Acoustic Neurinomas: Results, Time Course or Function and Rehabilitation, in M. Samii (Hrsg.): Skull Base Surgery, Springer, Hannover, 1992, S. 894–897.

[18] W. Yu, K. Daniilidis, G. Sommer: Rtated Wedge Averaging Method for Junction Characterization, in IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, USA, 1998, S. 390– 395.