Matthias Zobel, Arnd Gebhard, Dietrich Paulus, Joachim Denzler, Heinrich Niemann

**Robust Facial Feature Localization by Coupled Features**

appeared in:
Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG2000)
Grenoble, France
p. 2–7
2000

# Robust Facial Feature Localization by Coupled Features

M. Zobel, A. Gebhard, D. Paulus, J. Denzler, H. Niemann

Lehrstuhl für Mustererkennung (Informatik 5)
Universität Erlangen–Nürnberg
Martensstr. 3, 91058 Erlangen, Germany
{zobel,gebhard,paulus,denzler,niemann}@informatik.uni-erlangen.de

## Abstract

*In this paper, we consider the problem of robust localization of faces and some of their facial features. The task arises e.g. in the medical field of visual analysis of facial paresis. We detect faces and facial features by means of appropriate DCT coefficients that we obtain by neatly using the coding capabilities of a JPEG hardware compressor. Beside an anthropometric localization approach we focus on how spatial coupling of the facial features can be used to improve robustness of the localization. Because the presented approach is embedded in a completely probabilistic framework, it is not restricted to facial features, it can be generalized to multipart objects of any kind. Therefore the notion of a "coupled structure" is introduced. Finally, the approach is applied to the problem of localizing facial features in DCT–coded images and results from our experiments are shown.*

## 1. Motivation

The investigation of human faces by computer vision methods has become a major field of interest in the last decade [10, 8]. Two areas of research can be identified: detection and tracking of human faces, and feature detection in face images. The latter often depends on the first. Facial features are used for identification, access control, as well as in multi media applications. In particular, new coding and image transmission schemes depend on symbolic information on the faces to be sent. Facial features are also of interest in medical applications as we show in the following.

Modern multi media computers are equipped with special hardware for image compression. Whereas formerly the compressed domain was regarded unsuited for image analysis, since it suppresses high frequency information which is crucial for edge detection, lately image processing and image understanding is also partially done in the compressed domain.

In this paper we first describe how facial features can be easily detected in real time in JPEG encoded image sequences (Sect. 2). Then two approaches to localizing facial features are described. We start with an anthropometric based approach in Sect. 3 with no explicit representation of the relations between each of the facial features. A different, optimization based approach for localization of facial features using spatial dependencies between the features is presented in Sect. 4. A medical application is outlined in Sect. 5 and the proposals are validated by experiments which are described in Sect. 6.

## 2. DCT for Facial Feature Detection

In this section we shortly describe the Discrete Cosinus Transformation (DCT) and show how it can be used for the facial feature locating task.

Eq. (1) gives the calculation of 8×8 Forward DCT:

$$F(u,v) = \frac{1}{4}C(u)C(v)\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cdot \cos\frac{(2x+1)u\pi}{16}\cos\frac{(2x+1)v\pi}{16}, \quad (1)$$

$$\text{with} \quad C(u) = C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if} \quad u,v=0 \\ 1 & \text{else.} \end{cases}$$

For the localization of faces and facial features two kinds of information are used: on the one hand we use the *DC coefficient* (the coefficient $F(0,0)$ with no harmonic components) to get the average color value of an 8×8 block. This can, for example, be used to find face color regions in an

This work was supported by the "Deutsche Forschungsgemeinschaft" under grant SFB603/TP B2 and B3. Only the authors are responsible for the content.

image. The *AC coefficients* (all other DCT–coefficients containing harmonic components) include information about the gray value changes (or edge energy) inside a block. At a certain abstraction the facial features eyes and mouth produce certain horizontal lines into the image. The horizontal lines are represented by vertical edge energy. An approximation of the vertical edge energy denoted $b_v(j,k)$ inside block $(j,k)$ is the sum of coefficients (Fig. 1, upper right) of the first and second column of the DCT block.

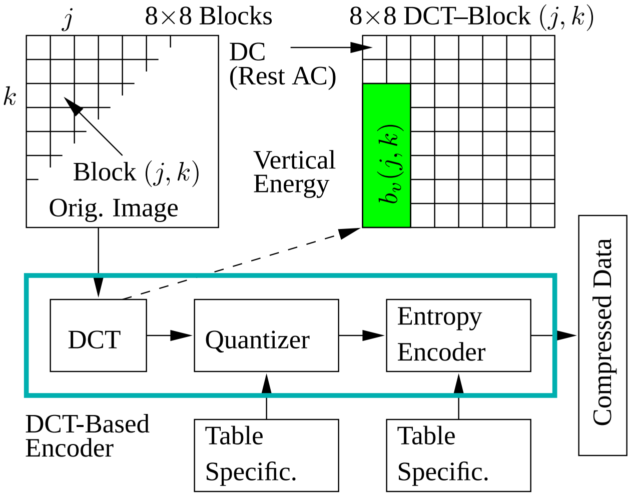

**Fig. 1. DCT–Based JPEG–Encoder**

Another thing makes DCT representation very interesting: it can be calculated on standard PCs in real time by means of a JPEG compressor that presently is a cheap hardware component. To obtain the desired DCT coefficients, one has to reverse the compression procedure (Fig. 1). The steps of the Entropy Encoder and the Quantizer are reversed in software and this can be performed in real time on a PC (for details cf. [10]).

In the next two sections we present two different approaches for the facial feature localization problem that both use the described DCT information.

## 3. Localization of Facial Features Using Anthropometric Knowledge

The first approach for facial feature localization is an extention of the localization method presented by Wang and Chang in [10]. In a first step a compact region with face color is found by means of the DC coefficients (for details cf. [10]).

The potential face region is then divided into 5 stripes (cf. Fig. 2, left). In the left halves to stripes 2 and 3 the right eye (region $\mathcal{A}_1$ in Fig. 2, right) is found. This is done by moving rectangular masks over the vertical edge energy representation (cf Sect. 2) of the image. The mask with the highest ratio of energy to area is considered to cover the eye. The left eye (region $\mathcal{A}_2$) is found with the optimal mask from region $\mathcal{A}_1$ in the right halves of the stripes 2 and 3. Analogically the mouth (region $\mathcal{A}_3$) is located by means of rectangular masks in the stripes 4 and 5.

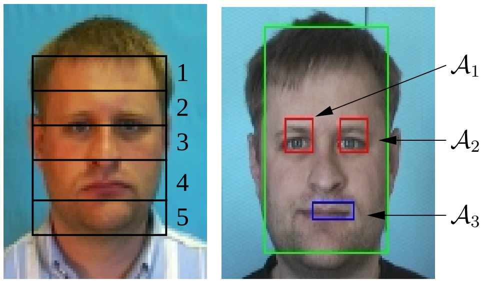

**Fig. 2. Left: Proportions of a human face. Right: Tracked face and facial features modeled as $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$**

Localization results are shown in Sect. 6.

## 4. Enhanced Coupled Localization

It was shown in the previous section that localization of the facial features eyes and mouth can be done by means of a physiologically motivated subdivision of the detected face region into five equally spaced stripes. The hypotheses of a face is rejected in a verification step, if, for example, one eye could not be located due to differences in position or energy.

To fix such false rejections the approach of Sect. 2 is improved. We show that the localization of features that have spatial relationships to each other can be done by solving an optimization problem. The main point is a *probabilistic* model that represents these spatial dependencies. For finding the locations of the features, one has to determine those parameters of the model that maximize the *a posteriori* probability (MAP) of the model conditioned by the current data.

The work which has mostly inspired us, is that of feature networks in [4]. There, the coupling of certain features as well as the composition of higher level geometric constraints is used to improve the accuracy of tracking. We focus on the coupling of features by means of a probabilistic model described in Sect. 4.1. In contrast to [4] we use a concrete model which is described in a probabilistic framework. It will be shown that the probabilistic model is strongly connected with the elastic, deformable contour extraction process by active contours [6]. In that framework an internal and an external energy exists. Such an elastic coupling of features by springs was introduced in [2] for facial feature tracking and later used in [12] in the context of deformable templates.

Our work combines the advantages of these approaches, simplifying the joint probabilistic data association filter (JPDAF) approach of [7] and reducing the whole estimation process to an energy minimization problem. Our approach can also be compared with active, elastic contours, where the contour points are substituted by higher level features; in our application, these features consists of the two eyes and the mouth (Sect. 4.3). The model parameters themselves can be estimated in a training step. In our current work, this is done by using a labeled training set. For this, the probabilistic framework is advantageous, because of the rich theory already available for parameter estimation, and the possibility of handling uncertainty, given by noisy data.

## 4.1. Coupled Structure – A Probabilistic Model

The model that is described in the following, is based on the active rays approach that is successfully used for contour based object tracking [1]. A 2D contour is represented by different 1D rays, which originate from one reference point. Instead of interpreting a point on a ray as a candidate for a contour point, it can be generally seen as the location of any given feature. So the concept of a contour in the image plane, which is represented by a given set of rays, is replaced by a general concept that we call *coupled structure*.

The position of a certain feature is given by a *coupled ray* $\boldsymbol{\varrho}_i = (\lambda_i, \phi_i)^T$ with length $\lambda_i$ and angle $\phi_i$. The pose of the ray is determined by the angle $\phi_i$ measured with respect to a given reference line in the image (usually the horizontal line). All coupled rays originate in a common point called the *coupling center* $\boldsymbol{m} = (m_x, m_y)^T$ with its image coordinates $m_x$ and $m_y$ (see Fig. 3). So the model, i.e. the coupled structure $\boldsymbol{s}$ is defined by

$$\boldsymbol{s} = (\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_n, \boldsymbol{m})^T. \tag{2}$$

Because of the fact that the locations of the features of the objects under consideration often change slightly (think of a non-rigid motion of a face) and that the detection of features is distorted by noise, it is reasonable to regard the important quantities of the model in a probabilistic way. This can be done by modeling the variations in the concrete values of the lengths $\lambda_i$ and angles $\phi_i$ of a ray $\boldsymbol{\varrho}_i$ by an appropriate probability density function

$$p_{\boldsymbol{\varrho}_i}(\lambda_i = l, \phi_i = \varphi | \boldsymbol{\varrho}_i). \tag{3}$$

By this representation it is intended to show explicitly the generality of the approach. For example, it can be thought of features that have more than one plausible location along a certain ray. So the necessity may arise to use multi-modal probability density functions. It is worth noting that the description can be extended to the 3–D case by using 3–D rays. Here, the description is restricted to features lying in one plane.

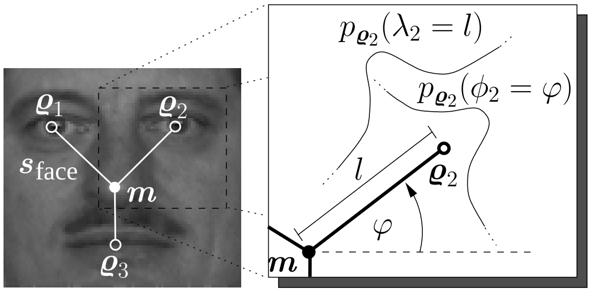

**Fig. 3. Modeling the spatial relations between facial feature using a coupled structure. The right side shows a magnification of one ray to explain the quantities.**

## 4.2. MAP Based Localization

We now treat the coupling structure $\boldsymbol{s}$ as a random vector in $\mathbb{R}^{2n+2}$. Then, a maximum a posteriori estimation for localizing the object can be applied. Spoken in different words, one has to seek for a parameter set $\boldsymbol{s}^* = (\boldsymbol{\varrho}_1, \ldots, \boldsymbol{\varrho}_n, \boldsymbol{m})^T$ which maximizes the posterior distribution $p(\boldsymbol{s}|\boldsymbol{f})$ conditioned on the image $\boldsymbol{f}$. Using Bayes' rule one gets

$$p(\boldsymbol{s}|\boldsymbol{f}) = \frac{p(\boldsymbol{f}|\boldsymbol{s})p(\boldsymbol{s})}{p(\boldsymbol{f})}, \tag{4}$$

where $p(\boldsymbol{f}|\boldsymbol{s})$ denotes the sensor model and $p(\boldsymbol{s})$ the prior of observing a certain configuration of our model. In a given reference coordinate system we can calculate $p(\boldsymbol{s})$ by

$$p(\boldsymbol{s}) = p(\boldsymbol{\varrho}_1) \cdot p(\boldsymbol{\varrho}_2) \cdot \ldots \cdot p(\boldsymbol{\varrho}_n) \cdot p(\boldsymbol{m}). \tag{5}$$

The independence assumption in (5) is valid, since the dependencies between different rays are implicitly given by the common coupling center $\boldsymbol{m}$. The joint probability $p(\boldsymbol{\varrho}_i) = p(\lambda_i|\phi_i)p(\phi_i)$ must be estimated from the data in the model generation process.

Now, for a transformation $\mathcal{T}$ of the model, for example, a rotation in the image plane, the corresponding density $p(\mathcal{T}\boldsymbol{s})$ is given by

$$p(\mathcal{T}\boldsymbol{s}) = |\det(J_{\mathcal{T}^{-1}}(\boldsymbol{s}))|\, p(\mathcal{T}^{-1}\boldsymbol{s}) \tag{6}$$

with $J_{\mathcal{T}^{-1}}$ being the Jacobian of the transformation $\mathcal{T}^{-1}$. A simple and useful transformation may be a global scaling operation, which influences only the length $\lambda_i$ of the ray $\boldsymbol{\varrho}_i$.

For the sensor model $p(\boldsymbol{f}|\boldsymbol{s})$ a common method is applied. We express the correspondence of the model $\boldsymbol{s}$ with the sensor data $\boldsymbol{f}$, i.e. the probability of observing $\boldsymbol{f}$ given the model, by a Gibbs (or Boltzmann) distribution of the form

$$p(\boldsymbol{f}|\boldsymbol{s}) = \frac{1}{z_{\text{ext}}} \exp\left[-E_{\text{ext}}(\boldsymbol{f}, \boldsymbol{s})\right] \tag{7}$$

with $z_{\text{ext}}$ being a normalizing constant. The term $E_{\text{ext}}(\boldsymbol{f}, \boldsymbol{s})$ can be interpreted as an external energy and needs to be specified dependent on the application. It should return high values for image data which do not correspond to the model, and low values for good matches.

Now, the estimation of the unknown parameter $\boldsymbol{s}^*$ can be described as an MAP estimation

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}}{\operatorname{argmax}} \frac{p(\boldsymbol{f}|\boldsymbol{s})p(\boldsymbol{s})}{p(\boldsymbol{f})}. \tag{8}$$

In the following subsection we give concrete examples of the model $\boldsymbol{s}$ in the area of localizing facial features as well as concrete terms for the prior $p(\boldsymbol{s})$ and the sensor model $p(\boldsymbol{f}|\boldsymbol{s})$.

### 4.3. Localizing Facial Features

For locating the facial features eyes and mouth, it is intuitive to model the spatial dependencies by a coupling structure $\boldsymbol{s}_{\text{face}}$ that consists of *three* coupling rays with the coupling center being the tip of the nose. There is one coupling ray for each eye and one for the mouth (cf. Fig. 3).

Since there is only one reasonable position for each facial feature in a face, the length and the angle of each ray are regarded as Gaussian distributed random variables, i.e.

$$p_{\boldsymbol{\varrho}_i}(\lambda_i = l) \propto \mathcal{N}({}^{\lambda}\mu_i, {}^{\lambda}\sigma_i^2), \text{ and} \tag{9}$$

$$p_{\boldsymbol{\varrho}_i}(\phi_i = \varphi) \propto \mathcal{N}({}^{\phi}\mu_i, {}^{\phi}\sigma_i^2). \tag{10}$$

Therefore it is sufficient to specify the means ${}^{l,\phi}\mu_i$ and variances ${}^{l,\phi}\sigma_i^2$ of this distributions for each ray $\boldsymbol{\varrho}_i$. They can be obtained, for example, by segmentation of a sample set of images taken from frontal views of different persons.

For the prior $p(\boldsymbol{s}_{\text{face}})$ in Eq. (5) it is necessary to specify explicitly $p(\boldsymbol{\varrho}_i)$. For the joint probability density function $p(\lambda_i, \phi_i)$ we write

$$p(\boldsymbol{\varrho}_i) = p(\lambda_i)p(\phi_i).$$

This independence assumption was verified by applying the $\chi^2$ test to data from 339 face images. Thus, we get for the prior $p(\boldsymbol{s}_{\text{face}})$ of our model parameters

$$p(\boldsymbol{s}_{\text{face}}) = p(\boldsymbol{m}) \prod_{i=1}^{3} p(\lambda_i)p(\phi_i). \tag{11}$$

Assuming a Gaussian distribution of the two parameters $\lambda_i$ and $\phi_i$ as mentioned earlier and a uniform distribution $p(\boldsymbol{m})$ over the image plane, i.e. no knowledge is used about the position of the face in the image, we get a distribution of the form

$$p(\boldsymbol{s}_{\text{face}}) = \frac{1}{z_{\text{int}}} \exp\left[-E_{\text{int}}(\boldsymbol{s}_{\text{face}})\right], \tag{12}$$

where $z_{\text{int}}$ is a normalizing constant and $E_{\text{int}}(\boldsymbol{s}_{\text{face}})$

$$E_{\text{int}}(\boldsymbol{s}_{\text{face}}) = \sum_{i=1}^{3} \frac{({}^{\lambda}\mu_i - \lambda_i)^2}{{}^{\lambda}\sigma_i^2} + \frac{({}^{\phi}\mu_i - \phi_i)^2}{{}^{\phi}\sigma_i^2}. \tag{13}$$

The term $E_{\text{int}}(\boldsymbol{s}_{\text{face}})$ can be interpreted as an internal energy of the model [9].

Thus this MAP approach can be seen as an energy minimization problem, with a term $E_{\text{int}}$ describing the deformation ability of the model and a second term $E_{\text{ext}}$ (cf. Eq. (7)) given by the image data conditioned on the model.

The external energy needs to be defined for the facial feature localization task. In Sect. 2 the features eyes and mouth are localized by means of vertical energies within DCT blocks. High vertical energies identify the unknown position of the facial features. Thus, it is natural to use this information for the coupled approach, too.

For each ray $\boldsymbol{\varrho}_i$ a certain rectangular area $\mathcal{A}_i(\boldsymbol{\varrho}_i)$ is defined, for which the vertical energies $b_v(j,k)$ of the DCT blocks are summed up (cf. Sect. 2). This results in an external energy for each ray $\boldsymbol{\varrho}_i$

$${}^{i}E_{\text{ext}} = \frac{1}{\sum\limits_{(j,k)\in\mathcal{A}_i(\boldsymbol{\varrho}_i)} b_v(j,k)} \tag{14}$$

that has high values for bad matches and low values for good ones. Finally, this leads to a total external energy of the coupled structure for the facial features

$$E_{\text{ext}}(\boldsymbol{f}, \boldsymbol{s}_{\text{face}}) = \sum_{i=1}^{3} {}^{i}E_{\text{ext}}. \tag{15}$$

With the prior of the model (5) and the sensor model (7) defined by the external energy (15) the unknown parameter set $\boldsymbol{s}^*_{\text{face}}$ can be determined using (8).

## 5. Application

Facial paresis is the most frequent paralysis which occurs isolated. At the Department of Otorhino–Larygology of our university over 100 patients with new appearances of paralysis are observed per year (cf. [11] for an overview).

The medical diagnosis of facial paresis bases on observations of a physician during the patient performs specific mimic movements. The diagnoses is generated by means of a medical index systems [5]. Thus a basic part of the diagnosis is performed on subjective human judgments. The goal of one of our projects is the development of a system for automatic diagnosis support and rehabilitation supervision of patients with facial paresis that also judges objectively by means of measurements inside the face.

A robust localization and tracking of the face and facial features is needed as the patient does not have to wear any artificial markers inside the face and additionally he is allowed to move slightly in front of the camera as the mimic exercises last for about one minute.

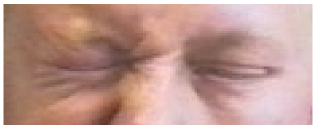

**Fig. 4. Eyes closing using maximal strength**

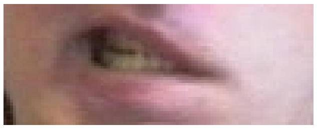

**Fig. 5. Teeth showing**

The symptoms of a facial paresis primarily appear in the regions of the eyes and mouth of the patient's face. In Fig. 4–Fig. 5 typical symptoms of facial paresis are shown for such regions. Fig. 4 shows the eye region of a patient who tries to close his eyes with maximal power. In the image both extreme asymmetries and functional deficits can be observed. That applies to the second image too, where the patient is performing the mimic exercise "Teeth Showing" (Fig. 5). These regions of interest are then subject to further feature detection; the features are used for an estimation for the degree of paralysis [3].

## 6. Experimental Results

For validation of the proposed approach, we performed experiments on a sample set of 335 faces of 20 different persons. In a first step, the positions of the two eyes and the mouth in all 335 images were achieved by hand–segmentation. Afterwards, the complete sample set was divided into a training part, which was used to estimate the parameters values of the reference coupled structure, and into a test part for evaluation.

To judge the quality of the results depending on the number of training examples, the 335 images were divided randomly into five sets of equal size. These five sets were used to perform five experiments with one set for training, and four sets for test, then 10 experiments with two sets for training and three sets for tests and so on. The whole procedure was done twice with different partitions of the five image sets. Thus, a total number of 10050 localizations were conducted.

The quality of facial feature extraction by the coupled structure was judged by computing the distances between the estimated position of the two eyes and the mouth, and the true position, obtained by the hand–segmentation. The best result using 335 training images is shown in Table 2. In Table 1 the comparable results are presented obtained by the method described in Sect. 2, i.e. without coupling.

| Results from 253 faces (82 faces not found) | | | | |
|---|---|---|---|---|
| | $\mu$ | $\sigma$ | min | max |
| Left eye | 1.61 | 1.67 | 0.00 | 13.86 |
| Right eye | 2.19 | 2.14 | 0.18 | 13.53 |
| Mouth | 2.14 | 2.12 | 0.18 | 16.60 |

**Table 1. Euclidean error for uncoupled localization. For each facial feature mean, standard deviation, minimal and maximal error in units of 8×8 blocks is given.**

| Results from 335 faces | | | | |
|---|---|---|---|---|
| | $\mu$ | $\sigma$ | min | max |
| Left Eye | 1.05 | 0.80 | 0.00 | 4.47 |
| Right Eye | 1.14 | 0.91 | 0.00 | 5.00 |
| Mouth | 1.69 | 1.20 | 0.00 | 6.71 |

**Table 2. Euclidean error for coupled localization. For each facial feature mean, standard deviation, minimal and maximal error in units of 8×8 blocks is given.**

The error between the estimated and the true position of the two eyes and the mouth could be reduced by $0.7$ blocks using the coupled approach. In total the mean error is below two blocks.

Finally, to show the robustness of the approach presented in Sect. 4, we prepared some images with artificial handmade distortions. The results can be seen in Fig. 6. In Fig. 6 (lower right), a result for localization of the facial features of one patient is shown.

## 7. Conclusion

We have described a method for localizing facial features in DCT coded images. This method has been improved by an approach of coupling such features in a probabilistic structure. Thus, spatial dependencies between multiple parts of objects are modeled. This leads to an improvement in the localization of the whole object in the case of distortions or wrong measurements and uncertainty in feature computation.

The experiments have shown, that an improvement from $2.0$ blocks error in the case of facial feature localization

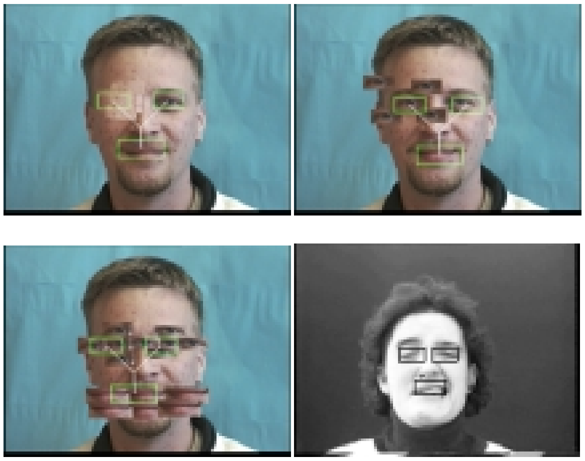

**Fig. 6. Results for artificially highly distorted face images: Left eye not present (upper left). More than one left eye visible (upper right), and more than one mouth as well as more than two eyes (lower left) are visible. Result of a patient image (lower right). The boxes mark the estimated position of the facial features.**

without modeling the spatial dependencies to $1.3$ blocks error using the coupling structure can be achieved. The advantage of the spatial modeling becomes obvious in the case of missing features (Fig. 6) due to occlusions or noisy data. The result itself is quite promising just because the features are simple and one can think of more sophisticated ones. We expect, that the mean error can be further reduced.

Summarizing the approach we like to emphasize that the idea of coupling different features of an object is natural and not new — as mentioned while giving the literature review in Sect. 4. Nevertheless, a complete formalization of this idea in a probabilistic framework, as given in the paper, has not be done until now. The main advantages arise from

1. the abstract description of the coupled structure, which will include 3–D objects in our future work; the position in 3–D can be estimated by integrating the transformation $\mathcal{T}$ (cf. Eq. 6) in the parameter estimation process (8).
2. the possibility to use multi–modal densities for describing the position of a certain feature,
3. the possibility to define different sensor models for each feature. In our case, this is demonstrated by the size of the rectangular area $\mathcal{A}_i(\boldsymbol{\varrho}_i)$, which differs between the two eyes and the mouth.

In our future work, we will focus on the integration of 3–D information, to handle rotating faces, too. There, we expect some problems with the computational effort in the practical realization of the MAP estimation by energy minimization. Additionally, we will apply more sophisticated sensor models to identify the facial features. Finally, the approach will be demonstrated on a different domain, to show its generality.

## References

[1] J. Denzler, B. Heigl, and H. Niemann. An efficient combination of 2d and 3d shape description for contour based tracking of moving objects. In H. Burkhardt and B. Neumann, editors, *Computer Vision - ECCV 98*, pages 843–857, Berlin, Heidelberg, New York, London, 1998. Lecture Notes in Computer Science.

[2] M. Fischler and R. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92, 1973.

[3] A. Gebhard, D. Paulus, and M. Dege. Grobe Lokalisation und Verfolgung von Patientengesichtern und Gesichtsmerkmalen in Farbbildfolgen in Echtzeit. In V. Rehrmann, editor, *Vierter Workshop Farbbildverarbeitung*, pages 97–103, Koblenz, 1998. Föhringer.

[4] G. Hager and K. Toyama. X vision: Combining image warping and geometric constraints for fast visual tracking. In A. Blake, editor, *Computer Vision - ECCV 96*, pages 507–517, Berlin, Heidelberg, New York, London, 1996. Lecture Notes in Computer Science.

[5] J. House. Facial nerve grading systems. *Laryngoscope*, 93:1056–1069, 1983.

[6] M. Kass, A. Wittkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 2(3):321–331, 1988.

[7] C. Rasmussen and G. Hager. Joint probabilistic techniques for tracking multi-part objects. In *Proceedings of Computer Vision and Pattern Recognition'98*, pages 16–21, 1998.

[8] M. Sanchez, J. Matas, and J. Kittler. Statistical chromaticity-based lip tracking with b-splines. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages 2973–2977, Munich, April 1997. IEEE Computer Society Press.

[9] D. Terzopoulos and R. Szeliski. Tracking with Kalman snakes. In A. Blake and A. Yuille, editors, *Active Vision*, pages 3–20. MIT Press, Cambridge, Massachusetts, London, England, 1992.

[10] H. Wang and S.-F. Chang. A Highly Efficient System for Automatic Face Region Detection in MPEG Video. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(4):615–628, August 1997.

[11] S. Wolf, M. Müller, W. Schneider, C. Haid, and M. Wigand. Facial nerve function after transtemporal removal of acoustic neurinomas: Results, time course or function and rehabilitation. In M. Samii, editor, *Skull Base Surgery*, pages 894–897. Hannover, 1992.

[12] A. Yuille and A. Blake. Deformable templates. In A. Blake and A. Yuille, editors, *Active Vision*, pages 21–38. MIT Press, Cambridge, Massachusetts, London, England, 1992.