EURASIP Journal on Applied Signal Processing 2004:6, 913–922

MAP Estimation of Chin and Cheek Contours in Video Sequences

Markus Kampmann
Ericsson Research, Ericsson Allee 1, 52134 Herzogenrath, Germany
Email: markus.kampmann@ericsson.com

Received 28 December 2002; Revised 8 September 2003

An algorithm for the estimation of chin and cheek contours in video sequences is proposed. This algorithm exploits a priori knowledge about shape and position of chin and cheek contours in images. Exploiting knowledge about the shape, a parametric 2D model representing chin and cheek contours is introduced. Exploiting knowledge about the position, a MAP estimator is developed taking into account the observed luminance gradient as well as a priori probabilities of chin and cheek contours positions. The proposed algorithm was tested with head and shoulder video sequences (image resolution CIF). In nearly 70% of all investigated video frames, a subjectively error free estimation could be achieved. The 2D estimate error is measured as on average between 2.4 and 2.9 pel.

Keywords and phrases: facial feature extraction, model-based video coding, parametric 2D model, face contour, face model.

1. INTRODUCTION

Techniques for estimation of facial features like eyes, mouth, nose, eyebrows, chin and cheek contours are essential for various types of applications [1, 2, 3, 4, 5, 6, 7, 8]. For facial recognition applications, features are estimated and used for recognition, authentification, and differentiation of human faces [7, 9, 10]. In multimedia data bases and information systems, facial feature estimation is required for analysis and indexing of human facial images. For specific video coding schemes like model-based video coding [11, 12, 13] (also sometimes called semantic video coding [14, 15] or object-based video coding [16, 17, 18]), facial feature estimation is also required. The estimated facial features are used for adaptation of a 3D face model to a person's face as well as for the determination of facial expressions [19, 20, 21, 22, 23].

In this paper, the estimation of chin and cheek contours is discussed. The estimation of chin and cheek is one of the most difficult tasks of facial feature estimation, especially that the chin contour is in many cases little visible. Furthermore, shadows, variations of the skin color, clothing, and double chin can complicate the estimation procedure. Rotations of the head (especially to the side) result in strong variations of the chin and cheek's shape and position. In this paper, head and shoulder video sequences are considered which are typical for news, videophone, or video conferencing sequences. Assuming a typical spatial resolution like the CIF format (352×288 luminance pels), the face size is quite small in those video sequences (with a typical face width from 40 to 70 pels). Taken this into account, the estimation of chin and cheek contours is further complicated.

In order to overcome these problems of chin and cheek contours estimation, the usage of a priori knowledge about these features is necessary. On one hand, knowledge about the typical shape of chin and cheek contours should be exploited. On the other hand, knowledge about more or less probable positions of chin and cheek contours should be taken into consideration.

In the literature, algorithms for chin and cheek contours estimation use a priori knowledge about shape and position only to a limited extent. Some approaches use edge detection or other basic image processing procedures for estimation [9]. Often, parametric 2D models (also called deformable templates [8]) for chin and cheek contours are exploited. Here, the model should be selected in such a way that an exact localization of the chin and cheek contours is possible. However, the number of unknown parameters should be as low as possible in order to increase the estimation's robustness. In [24, 25, 26], chin and cheek contours are approximated by ellipses resulting in quite large estimation errors. In [6, 21], parametric models consisting of two parabolas are used. A cost function is minimized to find the best fit of the parametric model to the chin. However, a two-parabola model is too rough for an exact representation of chin and cheek contours. For estimation, a person in the scene looking straight into the camera is assumed. No a priori knowledge about more or less probable positions of chin and cheek contours is exploited. In [22, 27], active contour models (*snakes*)

are used for the estimation of chin and cheek contours. A *snake* is an energy-minimizing spline influenced by image features to pull it toward edges. These approaches were applied to persons looking straight into the camera. Since the number of unknown parameters is high, the reliability of these algorithms is low [27].

In this paper, a new algorithm for chin and cheek contours' estimation is proposed. A priori knowledge about the typical shape and probable positions of chin and cheek contours is exploited in many ways. A new parametric 2D model representing chin and cheek contours is introduced. This 2D model consists of four parabola pieces which are linked together. The 2D model is described by eight parameters which have to be estimated. Assuming video sequences with a quite small face size, this model allows an exact localization of chin and cheek contours with a low number of parameters to be estimated. For estimation, a MAP estimator is developed. This estimator takes into account the observed luminance gradient as well as the probabilities of certain positions of chin and cheek contours. Besides, rotations of the head are also considered in the new estimator. For estimation, the positions of eyes and mouth are assumed to be known. In this paper, the algorithm from [20] is used for estimation of eyes and mouth middle positions.

The paper is organized as follows. In Section 2, the new parametric 2D model for chin and cheek contours is introduced. In Section 3, the chin contour is estimated, whereas the cheek contour is estimated in Section 4. Section 5 gives experimental results. A conclusion is given in Section 6.

2. PARAMETRIC 2D MODEL OF CHIN AND CHEEK CONTOURS

For representing the shape of chin and cheek contours, a parametric 2D model for these contours is introduced. The estimation of chin and cheek contours is done by estimation of the parameters of this 2D model. Figure 1 shows the parametric 2D model in a local, 2D system of coordinates (W, V). The origin of (W, V) lies in the middle of the intersection between the eyes middle points **r** and **l**. The W axis shows in the direction of the left eye middle point **l**. The 2D model consists of the four parts of a parabola P_1, P_2, P_3, and P_4 which are linked together. P_1 and P_2 represent the chin contour, while P_3 and P_4 the cheek contours. The endpoints $\mathbf{a} = (a_W, a_V)^T$ and $\mathbf{b} = (b_W, b_V)^T$ form the boundary of P_1, while the endpoints $\mathbf{a} = (a_W, a_V)^T$ and $\mathbf{c} = (c_W, c_V)^T$ form the boundary of P_2. A parabola part is unambiguously described by its two endpoints and the parabola axis. For the chin contour, the parabola axis A_0 is defined in such a way that A_0 is parallel to the V axis and **a** is a part of A_0. Therefore, P_1 and P_2 are completely described by the three endpoints $\mathbf{a} = (a_W, a_V)^T$, $\mathbf{b} = (b_W, b_V)^T$, and $\mathbf{c} = (c_W, c_V)^T$ only. So, six parameters have to be determined for the estimation of the chin contour.

The right cheek contour is described by the parabola piece P_3. The endpoints $\mathbf{b} = (b_W, b_V)^T$ and $\mathbf{d} = (d_W, d_V)^T$ form the boundary of P_3. For a complete description of P_3, its parabola axis A_3 is needed. A_3 can be constructed from the parameters of the chin contour. A_3 is defined in such a way that it passes the origin of (W, V) and divides chord s_{01} between **a** and **b** in the middle. Since the endpoints **a** and **b** are known after the chin contour estimation, only the position $\mathbf{d} = (d_W, d_V)^T$ is unknown for a complete description of P_3. **d** depends on another restriction. Cheek contours are often covered by hair and therefore impossible to estimate. So, **d** is defined in such a way that it passes the line S. S is parallel to the W axis with a distance L_C. L_C is chosen as $L_C = 0.15 L_{\mathrm{EM}}$ with the eye-mouth distance L_{EM} defined as the distance between the W axis and the mouth middle point **m**. So, only the W-coordinate d_W is necessary for a description of **d**. Corresponding to P_3, only the W-coordinate e_W is necessary for the description of P_4. Taken these two parameters for the cheek contours into account, eight parameters have to be estimated for the chin and cheek contours.

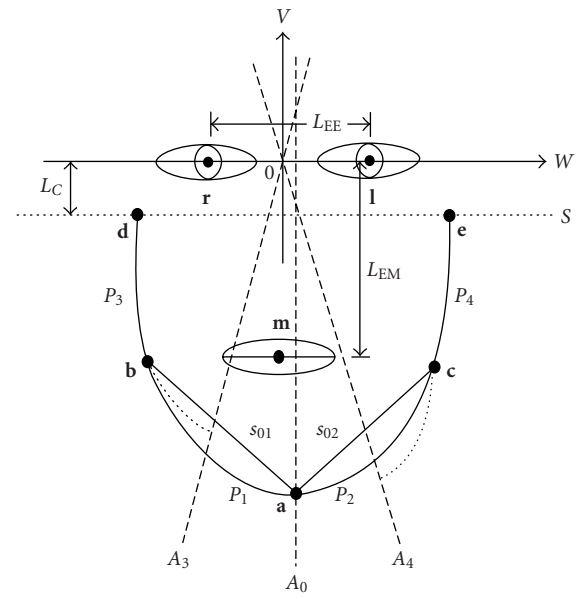

Figure 1: Parametric 2D model of chin and cheek contours consisting of four parabola pieces P_1, P_2, P_3, and P_4. **r** and **l** are the eyes middle points, and **m** the mouth middle point.

The estimation is carried out in two steps. First, the chin contour is estimated. Using the estimated chin contour, the cheek contours are estimated in a second step.

3. ESTIMATION OF CHIN CONTOUR

For estimation of the chin contour, the absolute value of the luminance gradient $|g(W, V)|$ is computed using the Sobel operator (Figure 2).

$|g(W, V)|$ is the observable measurement value that is used for estimation of the unknown parameters $\mathbf{a} = (a_W, a_V)^T$, $\mathbf{b} = (b_W, b_V)^T$, and $\mathbf{c} = (c_W, c_V)^T$. For simplification, these parameters are summarized to a parameter vector

(a) (b)

Figure 2: Luminance gradient: (a) luminance image; (b) absolute value of the luminance gradient determined by Sobel operator.

$\mathbf{f}_{\text{chin}} = (a_W, a_V, b_W, b_V, c_W, c_V)^T$. For chin contour estimation, an estimation algorithm is necessary which calculates an estimated value $\hat{\mathbf{f}}_{\text{chin}}$ from the known absolute value of the luminance gradient $|g(W,V)|$. Here, a MAP estimator is used. $\hat{\mathbf{f}}_{\text{chin}}$ is calculated using the MAP estimation algorithm according to

$$\hat{\mathbf{f}}_{\text{chin}} = \arg\max_{\mathbf{f}_{\text{chin}}} \{p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})\, p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}})\} \tag{1}$$

with $g = |g(W,V)|$. The conditional probability density function $p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})$ is called likelihood function, while $p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}})$ is the a priori probability density function of the parameter vector $\mathbf{f}_{\text{chin}}$. The product from likelihood function and a priori probability density function is called quality function. For calculation of $\hat{\mathbf{f}}_{\text{chin}}$, the quality function has to be established first. Then, the quality function is maximized by an optimization algorithm and the estimate value $\hat{\mathbf{f}}_{\text{chin}}$ is determined.

The likelihood function $p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})$ determines the probability for a measurement value g under the condition of a certain position $\mathbf{f}_{\text{chin}}$ of the chin contour. The determination of $p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})$ is difficult since manifold disturbances like shadows, clothing, or skin variations influence the observation g. Therefore, a simple approach is chosen in this work. Here, a proportional relation between $p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})$ and the mean absolute value of the luminance value along the chin contour is assumed:

$$p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}}) = c_{\text{chin}} \frac{1}{L_{P_1+P_2}} \int_{P_1+P_2} |g(W,V)|\, ds, \tag{2}$$

where $\int_{P_1+P_2} |g(W,V)| ds$ denotes the integral of the luminance gradient's absolute value along the parabola pieces P_1 and P_2; $L_{P_1+P_2}$ is the length of both parabola pieces; and c_{chin} a proportional constant. P_1 and P_2 are dependent on the parameters of $\mathbf{f}_{\text{chin}}$. According to (2), a high value of the mean luminance gradient corresponds to a high value of the likelihood function $p_{g|\mathbf{f}_{\text{chin}}}(g|\mathbf{f}_{\text{chin}})$. On the other hand, a low value means that the observed measurement belongs to the considered parameter vector with a low probability.

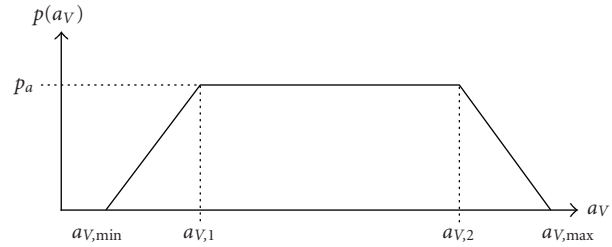

Figure 3: Probability density $p(a_V)$.

The probability density function $p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}})$ describes the probability of a certain chin contour position $\mathbf{f}_{\text{chin}} = (a_W, a_V, b_W, b_V, c_W, c_V)^T$. Due to the human anatomy, the bottom point **a** of the chin contour is located below the mouth and near the V axis. The W-coordinate a_W varies only slightly. The upper endpoints **b** and **c** are approximately located at the height of the mouth. The V coordinates b_V and c_V vary only little. Taking this into account, it is assumed that $p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}})$ is only dependent on the coordinates a_V, b_W, and c_W. Assuming a further independence between a_V on one side and b_W and c_W on the other side, $p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}})$ is equal to

$$p_{\mathbf{f}_{\text{chin}}}(\mathbf{f}_{\text{chin}}) = p(a_V)\, p(b_W, c_W). \tag{3}$$

First, $p(a_V)$ is examined. A range $a_{V,\min} < a_V < a_{V,\max}$ is set, whereas $a_{V,\min}$ and $a_{V,\max}$ are set proportional to the eye-mouth distance L_{EM} (see Figure 1). In case of talking, the mouth of a person is opened and closed. The position a_V is changing corresponding to the mouth movement. Due to the uniform movement, the probability $p(a_V)$ is not changed inside most part of the a_V range (Figure 3).

Therefore, $p(a_V)$ is set to

$$p(a_V) = \begin{cases} p_a \dfrac{a_V - a_{V,\min}}{a_{V,1} - a_{V,\min}}, & a_{V,\min} \le a_V \le a_{V,1}, \\ p_a, & a_{V,1} \le a_V \le a_{V,2}, \\ p_a \dfrac{a_V - a_{V,\max}}{a_{V,2} - a_{V,\max}}, & a_{V,2} \le a_V \le a_{V,\max}. \end{cases} \tag{4}$$

$p(a_V)$ is constant between $a_{V,1}$ and $a_{V,2}$. At the borders of the range, $p(a_V)$ is decreasing linearly. At $a_{V,\min}$ and $a_{V,\max}$, respectively, $p(a_V)$ is equal zero. $a_{V,1}$ and $a_{V,2}$ are set proportional to the eye-mouth distance L_{EM}.

Next, the term $p(b_W, c_W)$ in (3) is examined. First, ranges for b_W and c_W are introduced which are symmetrical to the V axis: $-b_{W,\max} < b_W < -b_{W,\min}$ and $b_{W,\min} < c_W < b_{W,\max}$. Here, $b_{W,\min}, b_{W,\max} > 0$ and are set proportional to the eye-eye distance L_{EE}. Since b_W and c_W are hardly influenced by the mouth movement, the assumption of a nearly uniform probability distribution is in contrast to $p(a_V)$ not useful. Considering instead that values of b_W, c_W have a higher probability in the middle of the corresponding range than at the borders, a sinus-like curve for $p(b_W)$ and $p(c_W)$

FIGURE 4: In case of a head rotation to the left side, a low value of $|c_W|$ corresponds to a high value of $|b_W|$.

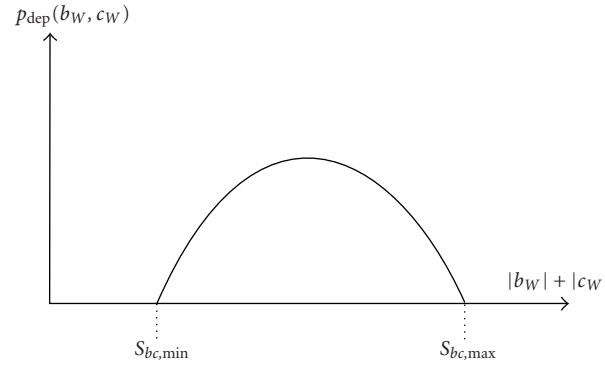

FIGURE 5: $p_{\text{dep}}(b_W, c_W)$ describes the dependence between b_W and c_W by the distance of the chin contour $|b_W| + |c_W|$.

is assumed:

$$p(b_W) = \frac{1}{2}\sin\left(\frac{b_W + b_{W,\max}}{b_{W,\max} - b_{W,\min}}\pi\right),$$
$$p(c_W) = \frac{1}{2}\sin\left(\frac{c_W - b_{W,\min}}{b_{W,\max} - b_{W,\min}}\pi\right). \tag{5}$$

In case of a statistical independence between b_W and c_W, the probability $p(b_W, c_W)$ could be expressed by

$$p(b_W, c_W) = p(b_W)p(c_W). \tag{6}$$

For this case, a certain value of c_W would have no influence on the occurrence of certain values of b_W. However, Figure 4 shows that a dependence between b_W and c_W exists.

In case of a head rotation to the left side, $|c_W|$ has a low value. In this case, $|b_W|$ has a high value. Therefore, an independence between b_W and c_W does not exist. In order to take their dependence into consideration, (6) is extended by an additional term $p_{\text{dep}}(b_W, c_W)$:

$$p(b_W, c_W) = p(b_W)p(c_W)p_{\text{dep}}(b_W, c_W). \tag{7}$$

According to Figure 4, a high value of $|b_W|$ corresponds to a low value of $|c_W|$ in case of a head rotation to the left side. In case of a head rotation to the right side, a low value of $|b_W|$ corresponds to a high value of $|c_W|$. Looking at the sum $|b_W| + |c_W|$ (which is the W distance of the chin contour endpoints), a middle value of $|b_W| + |c_W|$ is preferred in case of a head rotation. Low or high values of $|b_W| + |c_W|$ are less probable. According to this, $p_{\text{dep}}(b_W, c_W)$ is assumed to be

$$p_{\text{dep}}(b_W, c_W) = \frac{1}{2}\cos\left(\frac{|b_W| + |c_W| - s_{bc,\min}}{s_{bc,\max} - s_{bc,\min}}\pi\right) \tag{8}$$

in the range $s_{bc,\min} < |b_W| + |c_W| < s_{bc,\max}$ and

$$p_{\text{dep}}(b_W, c_W) = 0 \tag{9}$$

in all other areas (Figure 5).

The upper bound $s_{bc,\max}$ and the lower bound $s_{bc,\min}$ for the distance of the chin contour endpoints are set proportional to the eye-eye distance L_{EE}.

Using (2), (4), and (7), the quality function in (1) is completely known. The next step is the maximization of (1) and the determination of $\hat{\mathbf{f}}_{\text{chin}}$. The optimization is carried out in two steps. First, an initial value $\hat{\mathbf{f}}_{\text{chin,init}}$ is determined. Using $\hat{\mathbf{f}}_{\text{chin,init}}$, the final value $\hat{\mathbf{f}}_{\text{chin}}$ is determined in the second step. In the first step, search lines S_0, S_1, and S_2 are introduced (Figure 6). The initial values for the chin contour endpoints should be located on these lines. The lower search line S_0 for **a** is located on the V axis and is bounded by $a_{V,\min}$ and $a_{V,\max}$, respectively. The search lines S_1 and S_2 for **b** and **c** are on the height of the mouth middle point and parallel to the W axis. They are bounded by $-b_{W,\max}$, $-b_{W,\min}$ and $b_{W,\min}$, $b_{W,\max}$, respectively. Along these search lines, local maxima of $|g(W, V)|$ are determined. Only these local maxima could be the initial values for **a**, **b**, and **c**. For all combinations of these local maxima, the quality function in (1) is evaluated. The combination with the highest value of the quality function is chosen as initial estimate value $\hat{\mathbf{f}}_{\text{chin,init}}$. Taking $\hat{\mathbf{f}}_{\text{chin,init}}$ as a starting point, the final value $\hat{\mathbf{f}}_{\text{chin}}$ is determined in the following second step. 2D search areas are placed around the chin contour endpoints belonging to $\hat{\mathbf{f}}_{\text{chin,init}}$. Inside these search areas, the optimization is continued. Starting from the endpoints belonging to $\hat{\mathbf{f}}_{\text{chin,init}}$, the quality function in (1) is evaluated in an 8-point neighborhood around these endpoints. If the quality function is improved inside the 8-point neighborhood, the corresponding point is chosen as center for the next 8-point neighborhood evaluation. This procedure is continued until no more improvement of the quality function can be achieved. Then, the final estimate value $\hat{\mathbf{f}}_{\text{chin}}$ is found. The estimation of the chin contour is completed.

4. ESTIMATION OF CHEEK CONTOURS

Next, the cheek contours are estimated. The cheek contours are completely described by the parameter vector $\mathbf{f}_{\text{cheek}} = (d_W, e_W)^T$. The determination of the estimate value $\hat{\mathbf{f}}_{\text{cheek}}$ is

Figure 6: Search lines for initial estimation of the chin contour.

carried out analogous to the chin contour estimation. According to (1), a MAP estimator

$$\hat{\mathbf{f}}_{\text{cheek}} = \arg\max_{\mathbf{f}_{\text{cheek}}} \{p_{g|\mathbf{f}_{\text{cheek}}}(g|\mathbf{f}_{\text{cheek}})\, p_{\mathbf{f}_{\text{cheek}}}(\mathbf{f}_{\text{cheek}})\} \tag{10}$$

is introduced. Analogous to (2), $p_{g|\mathbf{f}_{\text{cheek}}}(g|\mathbf{f}_{\text{cheek}})$ is approximated by the integral over the absolute value of the luminance gradient along the parabola pieces P_3 and P_4:

$$p_{g|\mathbf{f}_{\text{cheek}}}(g|\mathbf{f}_{\text{cheek}}) = c_{\text{cheek}} \frac{1}{L_{P_3+P_4}} \int_{P_3+P_4} |g(W,V)|\, ds, \tag{11}$$

where $L_{P_3+P_4}$ denotes the length of both parabola pieces and c_{cheek} a proportional constant. $p_{\mathbf{f}_{\text{cheek}}}(\mathbf{f}_{\text{cheek}})$ is described by

$$p_{\mathbf{f}_{\text{cheek}}}(\mathbf{f}_{\text{cheek}}) = p(d_W)\, p(e_W)\, p_{\text{dep}}(d_W, e_W), \tag{12}$$

with, analogous to (5),

$$\begin{aligned} p(d_W) &= \frac{1}{2} \sin\left(\frac{d_W + d_{W,\max}}{d_{W,\max} - d_{W,\min}}\pi\right), \\ p(e_W) &= \frac{1}{2} \sin\left(\frac{e_W - d_{W,\min}}{d_{W,\max} - d_{W,\min}}\pi\right). \end{aligned} \tag{13}$$

According to (8), $p_{\text{dep}}(d_W, e_W)$ is described by

$$p_{\text{dep}}(d_W, e_W) = \frac{1}{2} \cos\left(\frac{|d_W| + |e_W| - s_{de,\min}}{s_{de,\max} - s_{de,\min}}\pi\right) \tag{14}$$

in the range $s_{de,\min} < |d_W| + |e_W| < s_{de,\max}$ and by

$$p_{\text{dep}}(d_W, e_W) = 0 \tag{15}$$

in all other areas.

Corresponding to $b_{W,\min}$, $b_{W,\max}$ and $s_{bc,\min}$, $s_{bc,\max}$, the values $d_{W,\min}$, $d_{W,\max}$ and $s_{de,\min}$, $s_{de,\max}$ are set proportional to the eye-eye distance L_{EE}. For determination of $\hat{\mathbf{f}}_{\text{cheek}}$, the search lines S_3, S_4 are introduced which are located on the line S (see Figure 1) and are bounded by $-d_{W,\max}$, $-d_{W,\min}$ and $d_{W,\min}$, $d_{W,\max}$, respectively. Along these search lines, local maxima of $|g(W,V)|$ are determined. Only these local maxima could be estimate values for d_W, e_W. For all combinations of these local maxima, the quality function in (10) is evaluated. The combination with the highest value of the quality function is the estimate value $\hat{\mathbf{f}}_{\text{cheek}}$. So, the estimation of the cheek contours is completed.

Table 1: Upper and lower bounds for chin and cheek parameters. L_{EE} denotes the distance between the eyes middle points, while L_{EM} denotes the distance between eyes and mouth.

Bound	Scale	Value
$a_{V,\min}$	L_{EM}	1.5
$a_{V,\max}$	L_{EM}	2.1
$a_{V,1}$	L_{EM}	1.6
$a_{V,2}$	L_{EM}	2.0
$b_{W,\min}$	L_{EE}	0.5
$b_{W,\max}$	L_{EE}	1.5
$d_{W,\min}$	L_{EE}	0.7
$d_{W,\max}$	L_{EE}	1.6
$s_{bc,\min}$	L_{EE}	1.6
$s_{bc,\max}$	L_{EE}	2.3
$s_{de,\min}$	L_{EE}	1.8
$s_{de,\max}$	L_{EE}	2.5

5. EXPERIMENTAL RESULTS

First, experiments were carried out in order to verify the assumed a priori probability density functions from Sections 3 and 4. Furthermore, upper and lower bounds for the probability density functions are determined.

In the second part, the proposed algorithm for chin and cheek contours estimation is tested with head and shoulder videophone sequences and its performance is evaluated.

5.1. Verification

For verification of the a priori probability density functions as well as for determination of the corresponding upper and lower bounds, tests were carried out. Here, 60 facial images (30 female and 30 male faces) from an image database were selected. The true positions of eyes and mouth middle positions and chin and cheek contours were manually determined from the facial images, and the parameters a_V, b_W, c_W, d_W, e_W, $|b_W| + |c_W|$, and $|d_W| + |e_W|$ were calculated. First, the upper and lower bounds $a_{V,\min}$, $a_{V,\max}$, $a_{V,1}$, $a_{V,2}$, $b_{W,\min}$, $b_{W,\max}$, $d_{W,\min}$, $d_{W,\max}$, $s_{bc,\min}$, $s_{bc,\max}$, $s_{de,\min}$, and $s_{de,\max}$ were determined. As described in Sections 3 and 4, $a_{V,\min}$, $a_{V,\max}$, $a_{V,1}$, and $a_{V,2}$ are set proportional to the eye-mouth distance L_{EM} and $b_{W,\min}$, $b_{W,\max}$, $d_{W,\min}$, $d_{W,\max}$, $s_{bc,\min}$, $s_{bc,\max}$, $s_{de,\min}$, and $s_{de,\max}$ are set proportional to the eye-eye distance L_{EE}. Table 1 shows the determined values for the upper and lower bounds extracted from the facial images.

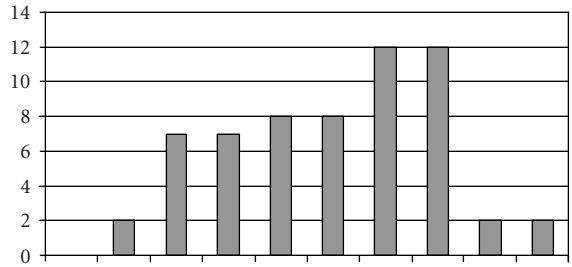

Figure 7: Frequency distribution for chin tip a_V. The value range $(a_{V,\min}, a_{V,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

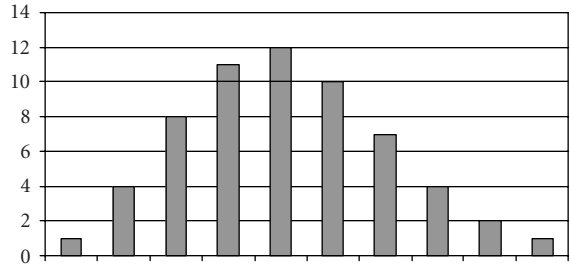

Figure 8: Frequency distribution for right chin contour endpoint b_W. The value range $(b_{W,\min}, b_{W,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

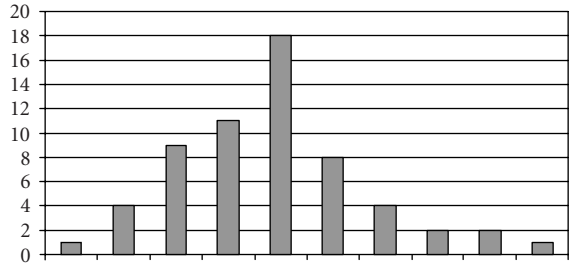

Figure 9: Frequency distribution for left chin contour endpoint c_W. The value range $(b_{W,\min}, b_{W,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

These values are used for the next step, the verification of the assumed a priori probability density functions from Sections 3 and 4. For all parameters a_V, b_W, c_W, d_W, e_W, $|b_W| + |c_W|$, and $|d_W| + |e_W|$, the corresponding frequency distribution using the 60 facial test images is calculated. Therefore, each parameter range is divided into 10 parts between its lower and upper bounds. For each part, the corresponding frequency of the parameter value within this part is determined. Figures 7, 8, 9, 10, 11, 12, and 13 show the results. For the chin tip position a_V, a uniform distribution was assumed in Section 3. For the other parameters, sinus-like distributions with more significant decreases towards the bounds were assumed. Looking at the frequency

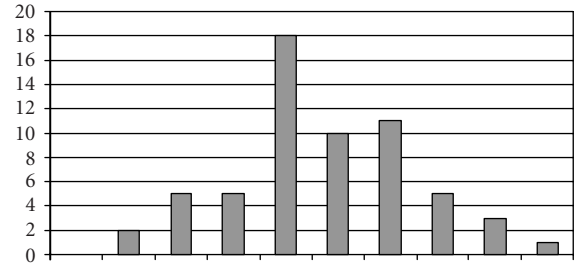

Figure 10: Frequency distribution for right cheek contour endpoint d_W. The value range $(d_{W,\min}, d_{W,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

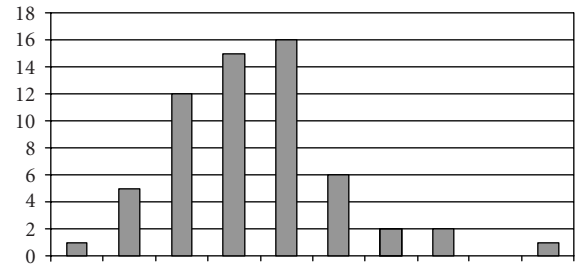

Figure 11: Frequency distribution for left cheek contour endpoint e_W. The value range $(d_{W,\min}, d_{W,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

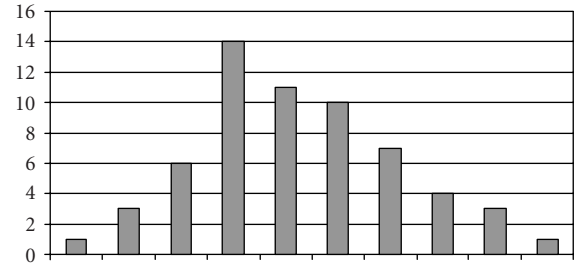

Figure 12: Frequency distribution for $|b_W|+|c_W|$ (distance between chin contour endpoints). The value range $(s_{bc,\min}, s_{bc,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

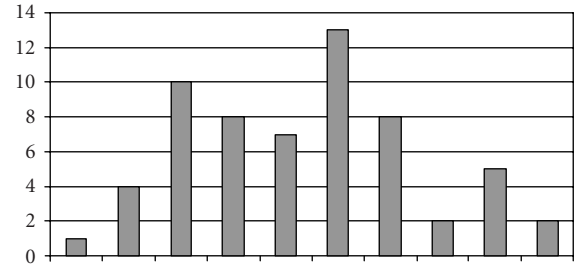

Figure 13: Frequency distribution for $|d_W|+|e_W|$ (distance between cheek contour endpoints). The value range $(s_{de,\min}, s_{de,\max})$ is subdivided into ten parts. For each part, the frequency out of 60 facial images is determined.

(a) (b) (c)

Figure 14: Test sequences: (a) Akiyo, (b) Miss America, and (c) Claire.

distributions from Figures 7, 8, 9, 10, 11, 12, and 13, these assumptions are verified in general. Whereas Figure 7 shows a more uniform distribution, the other figures show significant decreases towards the bounds.

However, further experiments with a larger number of facial test images should be carried out in the future in order to further check the assumed a priori probability density functions and the parameters' upper and lower bounds.

5.2. Performance evaluation

For evaluation of the proposed algorithm, the head and shoulder video sequences *Akiyo*, *Claire*, and *Miss America* with a resolution corresponding to CIF (352×288 luminance pels) and a frame rate of 10 Hz were used to test its performance (Figure 14). For the sequence Miss America, the person is mainly looking into the camera. For Claire, head rotation to the sides are observed. For Akiyo, the person is often looking down.

For evaluation of the algorithm's accuracy, the true positions of chin and cheek contours are manually determined from the video sequences. These true positions are then compared with the estimated ones to get the 2D estimate error in the image. Table 2 shows the estimate error's standard deviation for the test sequences. Here, it is distinguished between the chin tip **a**, the chin contour's upper points **b**, **c**, and the cheek contour's upper points **d**, **e**. Looking at the results, the estimate error for chin contour's upper points **b**, **c**, and cheek contour's upper points **d**, **e** are quite similar: 2.4 pel and 2.5 pel, respectively. The estimate error for the chin tip **a** is 2.9 pel, which is larger compared to the other four endpoints. The reason for this is mainly the video sequence Miss America, where the chin contour is very weak, disturbed by a shadow, and therefore difficult to estimate.

For additional evaluation, the estimation results are subjectively rated. In contrast to the results above, not only the positions of the five parabola pieces' endpoints are evaluated. Instead, the estimate of the complete chin and cheek contours is compared with the true ones. Three different subjective quality classes are introduced. In the first class, no deviation between the true and the estimated chin and cheek contours is observable, the estimation is error free. For the second quality class, an estimation error is observable. Finally, the third class means erroneous results, where the true contours are completely missed. For example, hair, clothing, lips, and so forth are detected instead of chin and cheek. All estimated chin and cheek contours are rated according to the three quality classes. Table 3 shows the achieved results. In nearly 70% of all frames, an error free estimation is possible. A completely missed estimation was observed in no frame.

Table 2: Standard deviation of 2D estimate errors for the chin and cheek contours (video sequences Akiyo, Claire, and Miss America).

Facial feature point	2D estimate error (pel)
Chin tip **a**	2.9
Chin contour's upper points **b**, **c**	2.4
Cheek contour's upper points **d**, **e**	2.5

Table 3: Percentage of estimated chin and cheek contours according to three quality classes (video sequences Akiyo, Claire, and Miss America).

Quality classes	Percentage (%)
(1) Error free	68
(2) Estimation error observable	32
(3) Complete mismatch	0

Figures 15, 16, 17, and 18 show examples of the estimated chin and cheek contours over the original images. Figures 15, 16, and 17 shows results of the first quality class with error free estimation. Results from the second quality class are given in Figure 18. Here, small deviations are noticed.

Since an accurate estimate of eyes and mouth middle positions is fundamental for the proposed chin and cheek estimation, an evaluation of the used algorithm from [20] for eyes and mouth estimation is given. Figures 15, 16, 17, and 18 show results for eyes and mouth middle positions estimation. A subjectively accurate estimation of eyes and mouth is observed. Measuring the estimate error for eyes and mouth in the same way as for chin and cheek, the estimate error's standard deviation is 1.5 pel for the eyes (here only open eyes are considered and the pupil position is taken as middle position) and 3.1 pel for the mouth.

Figure 15: Test sequence Akiyo: estimated chin and cheek contours over original images without estimation error (quality class 1). Displayed eyes and mouth middle positions are estimated by [20] and are known to the algorithm.

Figure 16: Test sequence Claire: estimated chin and cheek contours over original images without estimation error (quality class 1). Displayed eyes and mouth middle positions are estimated by [20] and are known to the algorithm.

6. CONCLUSIONS

A new algorithm for estimation of chin and cheek contours in video sequences is proposed. Within this algorithm, a priori knowledge about shape and position of chin and cheek contours is exploited. A parametric 2D model representing the shape of chin and cheek contours is introduced. This 2D model consists of four parabola pieces which are linked together. Eight parameters describe the parametric 2D model. Chin and cheek contours are estimated by determination of these eight parameters. Exploiting a priori knowledge about the position of chin and cheek contours, a MAP estimator is introduced. This MAP estimator takes into account the observed luminance gradient as well as a priori probabilities of the chin and cheek contours' positions. The estimation is done in two steps. First, the chin contour is estimated. In the second step, the cheek contours are determined.

Figure 17: Test sequence Miss America: estimated chin and cheek contours over original images without estimation error (quality class 1). Displayed eyes and mouth middle positions are estimated by [20] and are known to the algorithm.

Figure 18: Test sequences Akiyo, Claire, Miss America: estimated chin and cheek contours over original images with observable estimation errors (quality class 2). Displayed eyes and mouth middle positions are estimated by [20] and are known to the algorithm.

Using facial images from an image data base, the assumed a priori probabilities of the chin and cheek contours' positions were verified. Then, the proposed algorithm was tested with typical head and shoulders video sequences. In nearly 70% of all frames, a subjectively perfect estimation is possible. In no frame, a complete mismatch is noticeable. The standard deviation of the 2D estimate error is measured as 2.4 pel (upper endpoints of the chin contour), 2.5 pel (upper endpoints of the cheek contours), and 2.9 pel (chin tip), respectively.

A further advantage of the described algorithm is its flexibility. The assumed a priori probabilities could be easily exchanged by other functions if further measurements will suggest this.

ACKNOWLEDGMENT

This work has been carried out at the Institute of Communication Theory and Signal Processing, University of Hannover, Germany.

REFERENCES

[1] P. M. Antoszczyszyn, J. M. Hannah, and P. M. Grant, "Facial features motion analysis for wire-frame tracking in model-based moving image coding," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2669–2672, Munich, Germany, April 1997.

[2] G. Chow and X. Li, "Towards a system for automatic facial feature detection," *Pattern Recognition*, vol. 26, no. 12, pp. 1739–1755, 1993.

[3] I. Essa and A. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757–763, 1997.

[4] S.-H. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, "Facial feature detection using geometrical face model: an efficient approach," *Pattern Recognition*, vol. 31, no. 3, pp. 273–282, 1998.

[5] C. J. Kuo, R.-S. Huang, and T.-G. Lin, "3-D facial model estimation from single front-view facial image," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 183–192, 2002.

[6] M. J. T. Reinders, F. A. Odijk, J. C. A. van der Lubbe, and J. J. Gerbrands, "Tracking of global motion and facial expressions of a human face in image sequences," in *Proc. SPIE Visual Communications and Image Processing*, vol. 2904, pp. 1516–1527, Boston, Mass, USA, November 1993.

[7] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65–77, 1992.

[8] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.

[9] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[10] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–741, 1995.

[11] K. Aizawa and T. S. Huang, "Model-based image coding advanced video coding techniques for very low bit-rate applications," *Proceedings of the IEEE*, vol. 83, no. 2, pp. 259–271, 1995.

[12] C. S. Choi, K. Aizawa, H. Harashima, and T. Takebe, "Analysis and synthesis of facial image sequences in model-based image coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 4, no. 3, pp. 257–275, 1994.

[13] W. J. Welsh, S. Searby, and J. B. Waite, "Model-based image coding," *British Telecom Technology Journal*, vol. 8, no. 3, pp. 94–106, 1990.

[14] H. Musmann, "A layered coding system for very low bit rate video coding," *Signal Processing: Image Communication*, vol. 7, no. 4–6, pp. 267–278, 1995.

[15] L. Zhang, "Automatic adaptation of a face model using action units for semantic coding of videophone sequences," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 8, no. 6, pp. 781–795, 1998.

[16] M. Kampmann and J. Ostermann, "Automatic adaptation of a face model in a layered coder with an object-based analysis-synthesis layer and a knowledge-based layer," *Signal Processing: Image Communication*, vol. 9, no. 3, pp. 201–220, 1997.

[17] H. Musmann, M. Hötter, and J. Ostermann, "Object-oriented analysis-synthesis coding of moving images," *Signal Processing: Image Communication*, vol. 1, no. 2, pp. 117–138, 1989.

[18] J. Ostermann, "Object-based analysis-synthesis coding based on the source model of moving rigid 3D objects," *Signal Processing: Image Communication*, vol. 6, no. 2, pp. 143–161, 1994.

[19] P. M. Antoszczyszyn, J. M. Hannah, and P. M. Grant, "A comparison of detailed automatic wire-frame fitting methods," in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. 468–471, Santa Barbara, Calif, USA, October 1997.

[20] M. Kampmann, "Automatic 3-D face model adaptation for model-based coding of videophone sequences," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 3, pp. 172–182, 2002.

[21] M. J. T. Reinders, P. J. L. van Beek, B. Sankur, and J. C. van der Lubbe, "Facial feature location and adaptation of a generic face model for model-based coding," *Signal Processing: Image Communication*, vol. 7, no. 1, pp. 57–74, 1995.

[22] R. L. Rudianto and K. N. Ngan, "Automatic 3D wireframe model fitting to frontal facial image in model-based video coding," in *Proc. International Picture Coding Symposium (PCS '96)*, pp. 585–588, Melbourne, Australia, March 1996.

[23] Z. Wen, M. T. Chan, and T. S. Huang, "Face animation driven by contour-based visual tracking," in *Proc. International Picture Coding Symposium (PCS '01)*, pp. 263–266, Seoul, Korea, April 2001.

[24] H.-J. Lee, D.-G. Sim, and R.-H. Park, "Relaxation algorithm for detection of face outline and eye locations," in *Proc. IAPR Workshop on Machine Vision Applications*, pp. 527–530, Makuhari, Chiba, Japan, November 1998.

[25] E. Saber and A. M. Tekalp, "Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions," *Pattern Recognition Letters*, vol. 19, no. 8, pp. 669–680, 1998.

[26] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Processing: Image Communication*, vol. 12, no. 3, pp. 263–281, 1998.

[27] C.-L. Huang and C.-W. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, vol. 25, no. 12, pp. 1435–1444, 1992.

Markus Kampmann was born in Essen, Germany, in 1968. He received the Diploma degree in electrical engineering from the University of Bochum, Germany, in 1993, and the Doctoral degree in electrical engineering from the University of Hannover, Germany, in 2002. From 1993 to 2001, he was working as a Research Assistant at the Institute of Communication Theory and Signal Processing, the University of Hannover, Germany. His research interests were in the fields of video coding, facial animation, and image analysis. Since 2001, he is working with Ericsson Research in Herzogenrath, Germany. His working fields are multimedia streaming and mobile multimedia delivery.