Reprinted from

IMAGE COMMUNICATION

Signal Processing: *Image Communication* 10 (1997) 93–114

Tracking a face for knowledge-based coding of videophone sequences

Liang Zhang*

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Appelstraße 9A, D-30167 Hannover, Germany

kein pdf

ELSEVIER

face during the sequence using the face model (face tracking).

Recently, adaptation of the face model *Candide* onto the person's face in a knowledge-based coder has been investigated [9, 12, 21]. In these approaches, at the beginning of the sequence the person's face is first detected using template matching and feature extraction techniques. Then, the face model *Candide* is adapted to the size of the person's face and integrated into the 3D model object which describes the real object.

For face tracking during the sequence, several algorithms based on global head motion compensation under the assumption of a rigid face model have been proposed [5, 9, 14, 18, 23]. In [18, 23], motion compensation is based on extracting 2D correspondences, e.g. feature points or contours. Here, the locations of the feature points or contours have to be estimated first. Then, the 3D motion of the head is determined from these 2D correspondences in two successive frames under the assumption of a rigid face model. Hence, mismatching of the 2D correspondences directly affects the accuracy of face tracking. The other methods estimate the 3D head motion, i.e. rotation and translation, from two successive frames using the optical flow constraint equation [5, 9, 14]. Because the accuracy of the 3D motion estimation of the head is affected by the 3D model shape, the inaccurate 3D model shape reduces the accuracy of face tracking. In order to overcome the effects of inaccurate 3D model shape on the 3D motion estimation, several algorithms for the update of the face model's shape have been proposed [3–5, 13]. In [4, 5, 13], a depth update of the face model based on the optical flow constraint equation is proposed. In [3], the 3D motion and the 3D coordinates of the vertices of the face model are simultaneously estimated including photometric effects. Extensive overviews on this topic can be found in [1, 2, 6, 15, 22].

In this paper, an algorithm for tracking a face using the face model *Candide* for knowledge-based coding of videophone sequences is presented, which combines global head motion compensation and the update of the face model's shape during the sequence. For shape update, not only the eye and mouth center points of the face model are adapted to match the positions of the real eye and mouth in the sequence but also the orientation of the face model is updated. The proposed algorithm differs from the previous works in the following two aspects: (1) a flexible face model is used where the face model is adapted from frame to frame based on the estimated 3D eye and mouth center positions, in opposition to [9] where a rigid face model is used which is adapted only once at the beginning of the sequence; (2) compared to [3–5, 13], only the 3D eye and mouth center positions are estimated for the update of the face model's shape. The other vertices of the face model are moved based on these estimated 3D positions.

The paper is organized as follows. Section 2 briefly introduces the concept of a KBASC, the proposed face tracking system and the face model *Candide*. Section 3 reviews the global head motion compensation used in [9]. Section 4 discusses how the 2D eye and mouth center positions of the person's face in the image plane are estimated. Section 5 explains how from these 2D center positions first estimates of the 3D eye and mouth center positions can be derived. Section 6 reports how the final estimates of these 3D center positions are obtained by considering the face model's orientation. Experimental results with synthetic images and typical head-and-shoulder videophone sequences are given in Section 7.

2. KBASC, face tracking scheme and face model *Candide*

In this section, the concept of KBASC, the whole face tracking scheme, and the face model *Candide* used in this paper are addressed.

2.1. Concept of KBASC

KBASC tries to map a model world to the real world. The 3D *real objects* of the *real world* are captured by a *real camera*. This camera is modeled by a static pinhole *model camera* (Fig. 1), which looks into a 3D *model world*. In this model world, each moving real object is described by an opaque, diffusely reflecting moving 3D *model object*. The shape of the model object is represented by a 3D

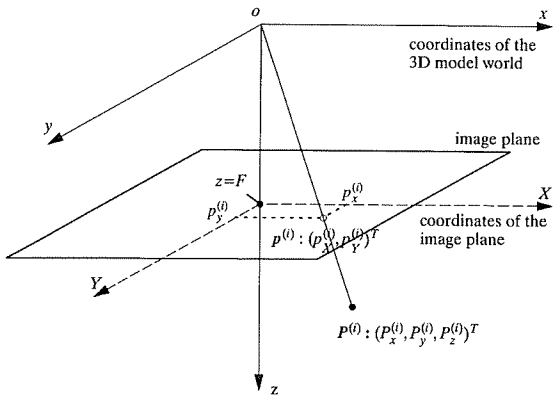


Fig. 1. Perspective camera model: F is the focal length of the camera.

wireframe (vertices) including predefined wireframe models. Each model object in the model world is described by three parameter sets defining the motion, shape and surface color of an object. The parameter sets of each model object are estimated by an image analysis. For head-and-shoulder scenes, KBASC [9] uses a predefined face model *Candide* to achieve a better modelling of the human face and exploits the knowledge of the face location for subjectively tuned bit allocation, so that the coding efficiency is improved compared to OBASC. A KBASC does not make use of the knowledge about the human facial expressions. If knowledge about facial expressions is available, it can be exploited by extending a KBASC to a semantic coding system [17].

2.2. Face tracking scheme

The proposed complete face tracking scheme consists of the following six steps (Fig. 2): the first step is tracking by global head motion compensation under the assumption of rigidity of the moving 3D objects as used in [9]; the second step gives the estimates of 2D center positions of the eyes and the mouth in the image plane; the third step deals with the estimation of the 3D center positions of the eyes and the mouth from the 2D eye and mouth center positions estimated in the second step; with these 3D estimates the face model's position and size is

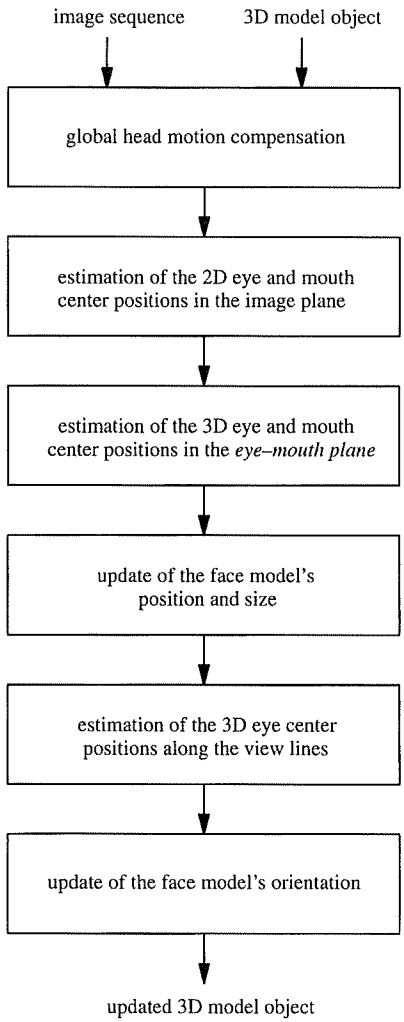


Fig. 2. Overview of the proposed complete face tracking system.

updated; the next step estimates the 3D eye center positions along the view lines; with these estimates finally the face model's orientation is updated. In this face tracking system, the first step gives a coarse face tracking and the other steps aim at a refined face tracking by updating the face model's shape.

2.3. Face model Candide

For the face model *Candide* (Fig. 3), its *shape parameters* are only the 3D center positions of the

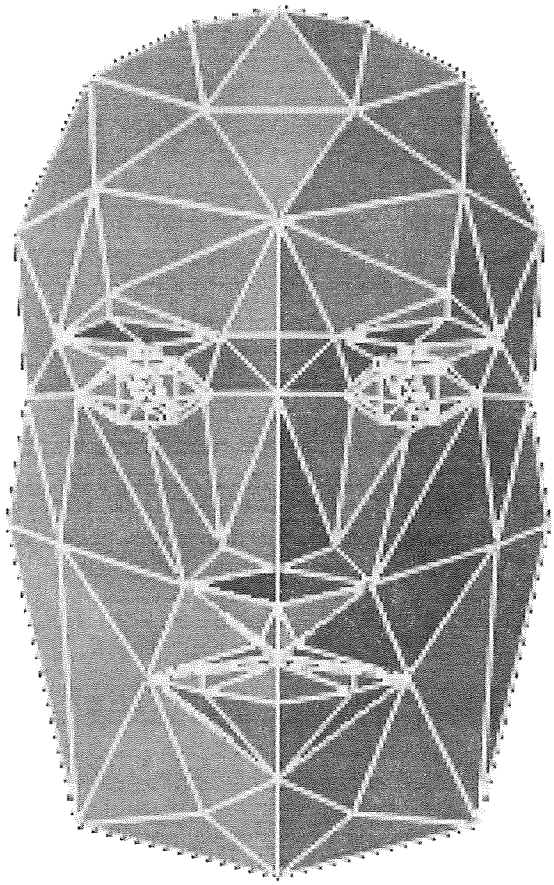


Fig. 3. Face model *Candide* [24].

eyes and the mouth, i.e. the position of the face model in the 3D world is determined only by these positions. In order to simplify the discussions in the following sections, two terminologies are introduced. The plane on which the 3D center positions of the eyes and the mouth of the face model lie is called *eye–mouth plane* and the norm of the *eye–mouth plane* is defined as the *orientation of the face model.*

Since the shape of the face model will be updated during the sequence, it is of interest how the vertices of the face model move. Let $\boldsymbol{P}^{*\prime}$, $\boldsymbol{P}^{*\mathrm{r}}$, $\boldsymbol{P}^{*\mathrm{m}}$ and $\boldsymbol{P}^{*\mathrm{o}}$ be the 3D positions of the left eye, the right eye, the mouth and the center between both eyes after an update of the face model's shape parameters, respectively, and let $\boldsymbol{P}^{\prime}$, $\boldsymbol{P}^{\mathrm{r}}$, $\boldsymbol{P}^{\mathrm{m}}$ and $\boldsymbol{P}^{\mathrm{o}}$ be those corresponding 3D positions before the update, respectively. An arbitrary point $\boldsymbol{P}^{(i)}$ on the surface of the face model moves to its new position $\boldsymbol{P}^{*(i)}$ as follows: at first, the face model is shifted with the translation vector $\boldsymbol{T}^{\mathrm{b}}$ so that the center $\boldsymbol{P}^{\mathrm{o}}$ between both eyes coincides with the origin of the 3D coordinate system; after this, the face model is rotated according to the 3D rotation angles, $\boldsymbol{R}_0^{\mathrm{b}} = (R_x^{\mathrm{b}}, R_y^{\mathrm{b}}, R_z^{\mathrm{b}})^{\mathrm{T}}$, scaled according to the scale factor $\boldsymbol{S} = (S_x, S_y, S_z)^{\mathrm{T}}$, and rotated according to the 3D rotation angles $\boldsymbol{R}_0^{\mathrm{a}} = (R_x^{\mathrm{a}}, R_y^{\mathrm{a}}, R_z^{\mathrm{a}})^{\mathrm{T}}$; finally, it is shifted with the translation vector $\boldsymbol{T}^{\mathrm{a}}$ to its new position. Thus, this motion can be formulated as

$$\boldsymbol{P}^{*(i)} = [\boldsymbol{R}_0^{\mathrm{a}}]\,[\boldsymbol{S}]\,[\boldsymbol{R}_0^{\mathrm{b}}]\,(\boldsymbol{P}^{(i)} - \boldsymbol{T}^{\mathrm{b}}) + \boldsymbol{T}^{\mathrm{a}}, \tag{2.1}$$

with the rotation matrices $[R_0^{\mathrm{b}}]$ and $[R_0^{\mathrm{a}}]$, and the scale factor matrix

$$[\boldsymbol{S}] = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & S_z \end{bmatrix}. \tag{2.2}$$

The superscripts a and b denote after and before scaling, respectively. The scale factors S_x, S_y and S_z along the x-, y- and z-axis are calculated from the *shape parameters* of the face model, i.e. 3D center positions of the eyes and the mouth,

$$S_x = \frac{\|\boldsymbol{P}^{*\prime} - \boldsymbol{P}^{*\mathrm{r}}\|}{\|\boldsymbol{P}^{\prime} - \boldsymbol{P}^{\mathrm{r}}\|}, \tag{2.3}$$

$$S_y = \frac{\|\boldsymbol{P}^{*\mathrm{m}} - \boldsymbol{P}^{*\mathrm{o}}\|}{\|\boldsymbol{P}^{\mathrm{m}} - \boldsymbol{P}^{\mathrm{o}}\|}, \tag{2.4}$$

$$S_z = \tfrac{1}{2}(S_x + S_y). \tag{2.5}$$

The rotation angles R_x^{b} and R_x^{a} are calculated as

$$R_x^{\mathrm{b}} = -\arctan\left\{\frac{P_z^{\mathrm{m}} - P_z^{\mathrm{o}}}{P_y^{\mathrm{m}} - P_y^{\mathrm{o}}}\right\}, \tag{2.6}$$

$$R_x^{\mathrm{a}} = \arctan\left\{\frac{P_z^{*\mathrm{m}} - P_z^{*\mathrm{o}}}{P_y^{*\mathrm{m}} - P_y^{*\mathrm{o}}}\right\}. \tag{2.7}$$

In the same way, the other rotation angles R_y^{b}, R_z^{b} and R_y^{a}, R_z^{a} can be calculated from the 3D eye and mouth center positions.

3. Global head motion compensation

As a first step of the algorithm proposed in this paper, the method of face tracking by global head motion compensation based on 3D rigid moving objects as used in [9] is applied. This method is reviewed in this section.

In [9], the face location and the silhouette of the person are exploited to generate a model object with head and shoulders (Fig. 4). The head includes the face model *Candide*. The knowledge about the face location is used to control the subdivision of the model object into head and shoulders. It is shown in [16] for OBASC that this subdivision can improve the 3D motion compensation especially when head and shoulders move differently. Furthermore, the face model matches the shape of a real person's face better than the generalized cylinder used by OBASC [16, 19]. Thus, 3D motion compensation of a KBASC should be superior to that of an OBASC [9].

3D motion estimation minimizes the mean square luminance difference between the model image and the real image. It is assumed that object components are rigid and have diffusely reflecting surfaces. Furthermore, diffuse illumination of the scene is assumed. Hence, color parameters are temporarily constant. Based on these assumptions, the luminance differences between two consecutive images s_k and s_{k-1} are due to object motion. For minimization of the luminance differences, an approach with a linearized signal model was developed in [10, 11] and further improved in [19]. According to [25], this method is not much sensitive to illumination effects. The luminance differences at position $\boldsymbol{P}^{(j)} = (P_x^{(j)}, P_y^{(j)}, P_z^{(j)})$ of the observation point $\boldsymbol{O}^{(j)} = (\boldsymbol{P}^{(j)}, \boldsymbol{g}^{(j)}, \boldsymbol{I}^{(j)})$ on the head model surface with N control points (vertices) $\boldsymbol{P}_C^{(i)}$ is related to motion parameters by the following linearized equation where the superscripts (j) are omitted for better readability [11]:

$$\begin{aligned}\Delta I = {} & F\, g_x/P_z\, T_x^{\mathrm{h}} + F\, g_y/P_z\, T_y^{\mathrm{h}} \\ & - [(P_x g_x + P_y g_y) F/P_z^2 + \Delta I/P_z] T_z^{\mathrm{h}} \\ & - [[P_x g_x (P_y - C_y^{\mathrm{h}}) + P_y g_y (P_y - C_y^{\mathrm{h}}) \\ & + P_z g_y (P_z - C_z^{\mathrm{h}})] F/P_z^2 + \Delta I/P_z (P_y - C_y^{\mathrm{h}})] R_x^{\mathrm{h}} \\ & + [[P_y g_y (P_x - C_x^{\mathrm{h}}) + P_x g_x (P_x - C_x^{\mathrm{h}}) \\ & + P_z g_x (P_z - C_z^{\mathrm{h}})] F/P_z^2 + \Delta I/P_z (P_x - C_x^{\mathrm{h}})] R_y^{\mathrm{h}} \\ & - [g_x (P_y - C_y^{\mathrm{h}}) - g_y (P_x - C_x^{\mathrm{h}})] F/P_z\, R_z^{\mathrm{h}},\end{aligned} \tag{3.1}$$

with the unknown motion parameters of the head model $\boldsymbol{T}^{\mathrm{h}} = (T_x^{\mathrm{h}}, T_y^{\mathrm{h}}, T_z^{\mathrm{h}})^{\mathrm{T}}$ and $\boldsymbol{R}_{C^{\mathrm{h}}} = (R_x^{\mathrm{h}}, R_y^{\mathrm{h}}, R_z^{\mathrm{h}})^{\mathrm{T}}$,

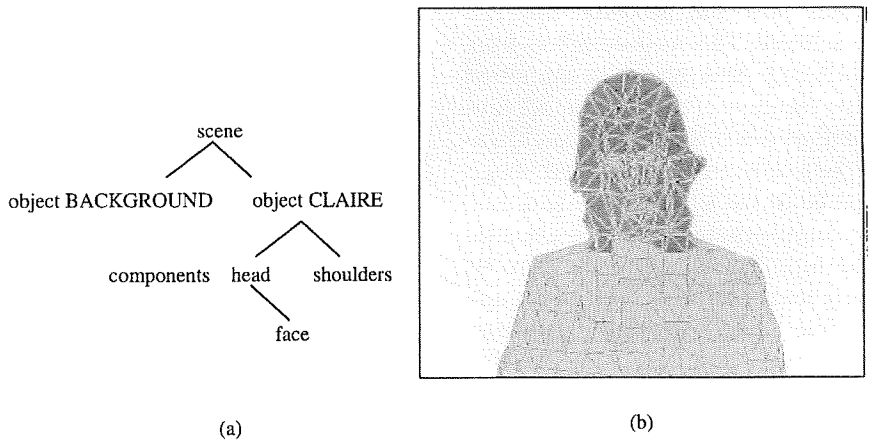


Fig. 4. Model scene and model object CLAIRE subdivided into the two flexibly connected components head and shoulders. The component head contains the face model *Candide*: (a) scene consisting of two objects; (b) components of model object CLAIRE.

the head model center $\boldsymbol{C}^{\mathrm{h}} = (C_x^{\mathrm{h}}, C_y^{\mathrm{h}}, C_z^{\mathrm{h}}) = (1/N) \times \sum_{i=1}^{N} \boldsymbol{P}_C^{(i)}$ and the linear gradients $\boldsymbol{g}^{(j)} = (g_x^{(j)}, g_y^{(j)})^T$ at position $\boldsymbol{P}^{(j)} = (P_x^{(j)}, P_y^{(j)}, P_z^{(j)})^{\mathrm{T}}$. The residuum of this equation system is minimized by linear regression:

$$\sum_{O(j)} (\Delta I^{(j)})^2 \rightarrow \mathrm{MIN}. \tag{3.2}$$

Due to the linearization, the motion parameters have to be estimated iteratively. After that, the motion of the face model *Candide* is compensated.

4. Estimation of the 2D center positions of the eyes and the mouth in the image plane

For estimation of the 2D center positions of the eyes and the mouth, an approach in [9, 12, 21] is presented only for adaptation of the face model once at the beginning of the sequence. Here, this approach is also used for tracking a face during the sequence and extended in several respects in order to achieve a reliable and more precise estimation of the eye and mouth center positions. The elements of the improved method are discussed in this section.

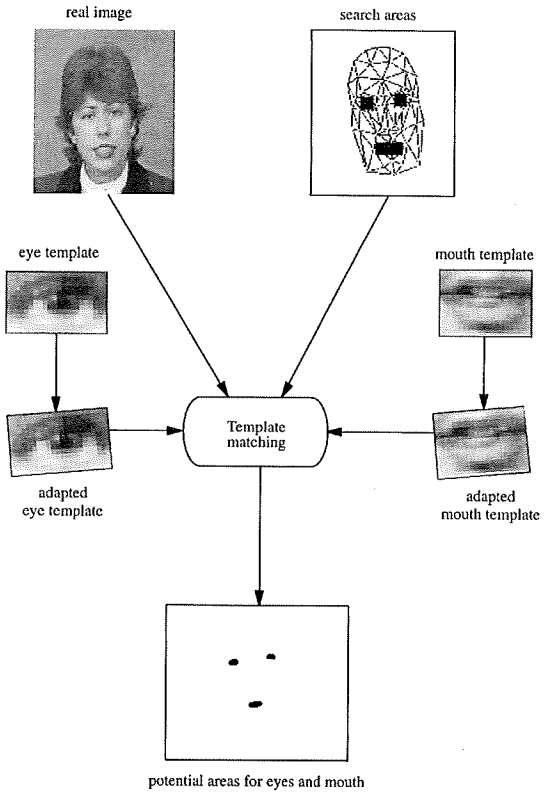


Fig. 5. Selection of the potential eye and mouth areas.

4.1. Selection of the potential eye and mouth areas

Fig. 5 illustrates how the potential eye and mouth areas are selected. This method differs from that used in [9, 12, 21]. After global head motion compensation, the location of the face model projected onto the image plane is already a good approximation of the face location in the real image. Therefore, compared to the large search areas used in [9, 12, 21], for simplicity smaller search areas for the eyes and the mouth are determined here by exploiting the projected positions of the eyes and the mouth of the face model onto the image plane. By this, the correct center positions of the eyes and the mouth can be found more often and the search areas can be essentially diminished. After that, in order to further reduce the search areas for the eyes and the mouth, template matching is used in which the eye and mouth templates are the same as those used in [9, 21]. The eye template is an open eye and has 23×13 picture elements. The mouth template has 31×21 picture elements. These templates are then roughly adapted to the size of the person's face. In contrast to [9, 12, 21], here the templates are additionally inclined in accordance with the inclination of the head. The inclination of the person's head is defined as

$$\beta = \arctan\left(\frac{Y_\ell - Y_\mathrm{r}}{X_\ell - X_\mathrm{r}}\right), \tag{4.1}$$

where (X_ℓ, Y_ℓ) and $(X_\mathrm{r}, Y_\mathrm{r})$ are the projections of the left eye and the right eye of the face model after global head motion compensation onto the image plane. For a point (X, Y) in the eye search areas, the correlation coefficient $c_{\mathrm{eye}}^k(X, Y)$ between the real image s_k and an eye template t_{eye} is computed in a window centered at the corresponding point

(X, Y) in the real image s_k. The correlation coefficient $c^k_{\text{eye}}(X, Y)$ is defined as follows:

$$c^k_{\text{eye}}(X, Y) = \frac{E(s_k\, t_{\text{eye}}) - m_{s_k} m_{t_{\text{eye}}}}{\sigma_{s_k}\ \sigma_{t_{\text{eye}}}} \tag{4.2}$$

where $E(\ )$ is a mean operation, $m_{t_{\text{eye}}}$, $\sigma^2_{t_{\text{eye}}}$ and m_{s_k}, $\sigma^2_{s_k}$ are the means and variances of the eye template t_{eye} and the real image s_k in the window which is as large as the eye template, respectively. The higher the value $c^k_{\text{eye}}(X, Y)$, the higher is the probability that this point (X, Y) in the eye search areas is the 2D eye center point. Points with high values of $c^k_{\text{eye}}(X, Y)$ are extracted as the potential areas for eyes. By this, the search areas for eyes are further reduced. In the same way, the mouth search area can be also reduced. The subsequent algorithms for estimation of eye and mouth center positions are applied only to these potential areas which might contain the eyes and the mouth.

4.2. Estimation of the 2D eye center positions

The pupil is defined as the center of the eye. Because the pupil of an eye is darker than the rest of the eye, a measure f_{eye} is evaluated within the potential areas for eyes:

$$f_{\text{eye}} = \frac{255 - s_k(X, Y)}{255}. \tag{4.3}$$

This measure assigns a high value to dark pels (X, Y) of the real image s_k. Those two points with the highest values of f_{eye} are selected as 2D eye center positions.

4.3. Estimation of the 2D mouth center position

For estimation of the mouth center position, a measure f_{mouth} is evaluated:

$$f_{\text{mouth}} = c^k_{\text{mouth, size}}(X, Y), \tag{4.4}$$

where $c^k_{\text{mouth,size}}(X, Y)$ are the correlation coefficients between the real image s_k and a mouth template. In contrast to Section 4.1, additional scaled templates with scale factors 0.8, 0.9, ..., 1.4 are evaluated. The point with the highest values of f_{mouth} is selected as 2D mouth center position.

5. Estimation of the 3D eye and mouth center positions in the *eye–mouth plane* and update of the face model's position and size

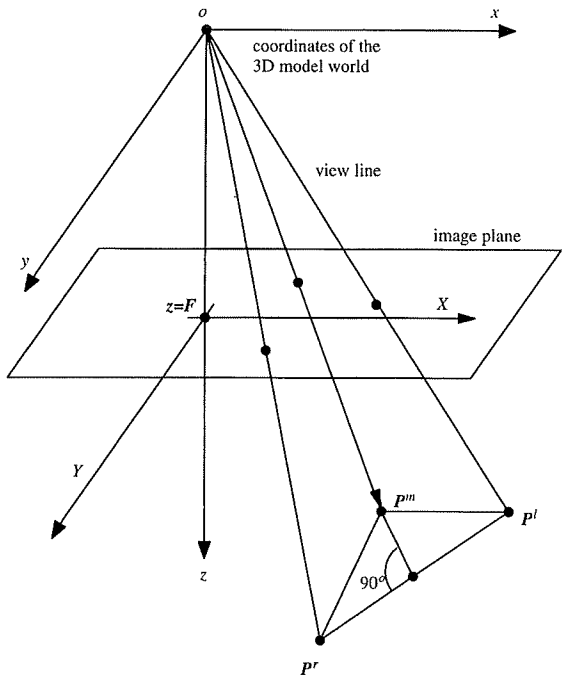
After the 2D center positions of the eyes and the mouth have been estimated as described in Section 4, the 3D center positions in the *eye–mouth plane* can be derived [26]. If the 2D center positions of the eyes could not at all be estimated, e.g. if the eyes are closed, no update of the face model's position and size is performed. In these situations, only global head motion compensation is applied for tracking the face. For the update of the face model's position and size during the sequence, the new 3D eye and mouth center positions of the face model have to be calculated by means of those 2D center positions just estimated. This problem is addressed in the following sections.

5.1. Overview of the calculation of the 3D eye and mouth center positions of the face model Candide

Since the true depths of the eyes and the mouth are unknown, the calculation of the new 3D center positions of the eyes and the mouth is based on the following two assumptions:

1. 3D center positions of the eyes and the mouth are in the same *eye–mouth plane* before and after the update of the face model's position and size;
2. 3D center positions of the eyes and the mouth span an isosceles triangle (Fig. 6) in the *eye–mouth plane.*

The projections of the 2D center positions of the eyes and the mouth estimated as described in Section 4 from the image plane onto the *eye–mouth plane* of the face model may not fulfill the second assumption. Because of a more precise estimation of the eye positions than of the mouth position in the image plane, the new 3D eye center positions of the face model will be calculated first. After that, the new 3D mouth center positions of the face model are calculated according to the second assumption.



Isosceles triangle assumption:

$$\| P^m - P^r \| = \| P^m - P^l \|$$

Fig. 6. Isosceles triangle assumption.

5.2. Calculation of the new 3D eye center positions of the face model

The projections of the estimated 2D eye center positions from the image plane onto the *eye–mouth plane* are taken as the new 3D eye center positions. According to the first assumption, the *eye–mouth plane* can be determined as follows:

$$\begin{vmatrix} P_x - P_x^m & P_y - P_y^m & P_z - P_z^m \\ P_x' - P_x^m & P_y' - P_y^m & P_z' - P_z^m \\ P_x^r - P_x^m & P_y^r - P_y^m & P_z^r - P_z^m \end{vmatrix} = 0, \tag{5.1}$$

where $\boldsymbol{P}^m = (P_x^m, P_y^m, P_z^m)$, $\boldsymbol{P}' = (P_x', P_y', P_z')$ and $\boldsymbol{P}^r = (P_x^r, P_y^r, P_z^r)$ are the 3D center positions of the mouth, the left eye and the right eye of the face model after global head motion compensation, respectively. $\boldsymbol{P} = (P_x, P_y, P_z)$ is an arbitrary point in the *eye–mouth plane*. The new 3D eye center positions $\boldsymbol{P}'' = (P_x'', P_y'', P_z'')$ and $\boldsymbol{P}'^r = (P_x'^r, P_y'^r, P_z'^r)$ to be calculated must satisfy a perspective projection of a static pinhole camera,

$$P_x'' = \frac{X_l}{F} P_z'', \qquad P_y'' = \frac{Y_l}{F} P_z'' \tag{5.2}$$

and

$$P_x'^r = \frac{X_r}{F} P_z'^r, \qquad P_y'^r = \frac{Y_r}{F} P_z'^r, \tag{5.3}$$

where F is the focal length of the *model camera*, (X_l, Y_l) and (X_r, Y_r) are the estimated 2D center positions of the left eye and the right eye. The value of F is known in the *model camera*. Substituting (5.2) into (5.1) gives the z-coordinate of the left eye

$$P_z'' = F \frac{P_x^m J_{11} + P_y^m J_{12} + P_z^m J_{13}}{X_l J_{11} + Y_l J_{12} + F J_{13}}, \tag{5.4}$$

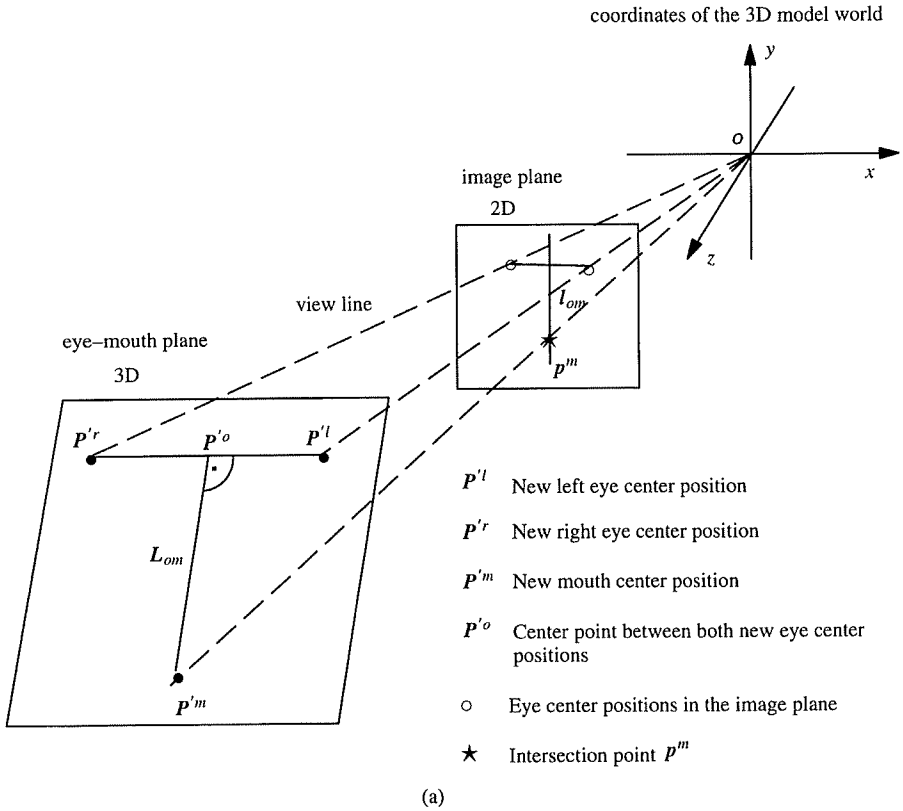
where J_{11}, J_{12}, and J_{13} are the subdeterminants of the determinant in (5.1). After P_z'' is calculated, P_x'' and P_y'' of the left eye can be derived from (5.2). Also, the z-coordinate $P_z'^r$ of the right eye can be similarly obtained by substituting (5.3) into (5.1), i.e.

$$P_z'^r = F \frac{P_x^m J_{11} + P_y^m J_{12} + P_z^m J_{13}}{X_r J_{11} + Y_r J_{12} + F J_{13}}. \tag{5.5}$$

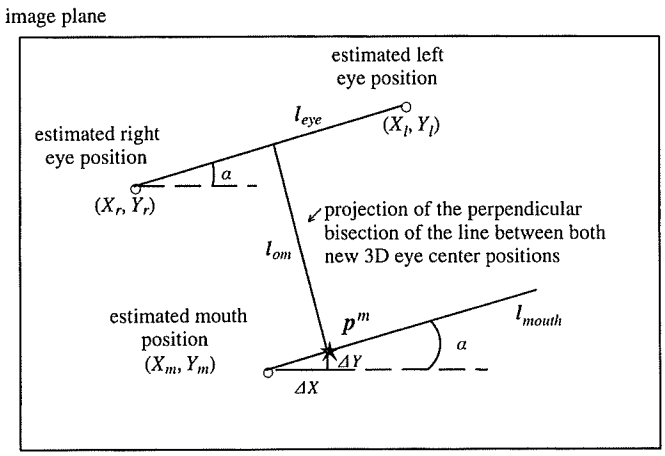
Substituting $P_z'^r$ into (5.3) gives the coordinates $P_x'^r$ and $P_y'^r$ of the right eye.

5.3. Calculation of the new 3D mouth center positions of the face model

Calculation of the new 3D mouth center positions of the face model is more complex than that of the eyes. According to the second assumption, the new 3D mouth center position $\boldsymbol{P}'^m = (P_x'^m, P_y'^m, P_z'^m)$ to be calculated lies on the perpendicular bisection $\boldsymbol{L}_{om}$ of the line $\boldsymbol{P}''\boldsymbol{P}'^r$ between both new 3D eye positions (Fig. 7(a)). This line $\boldsymbol{L}_{om}$ is projected onto the image plane. Here, the projected line l_{om} intersects a parallel l_{mouth} of the line l_{eye} which connects both estimated 2D eye positions in the image plane (Fig. 7(b)). This parallel l_{mouth} goes through the estimated 2D mouth position in the image plane. The projection of this intersection point $\boldsymbol{p}^m$ from the image plane onto the *eye–mouth plane* gives the new 3D mouth center positions $\boldsymbol{P}'^m$ (Fig. 7(a)).



(a)



(b)

Fig. 7. Calculation of the new 3D mouth center position of the face model. (a) Calculation of the perpendicular bisection of $P'^l P'^r$ and projection onto the image plane; projection of p^m onto the *eye–mouth plane*. (b) Update of the estimated mouth center position.

Due to the first assumption, the calculated 3D mouth center positions must satisfy (5.1) and due to the second assumption, the equation

$$(P'^{\prime\prime}_x - P'^{\mathrm{r}}_x)(P'^{\mathrm{m}}_x - P'^{\mathrm{o}}_x) + (P'^{\prime\prime}_y - P'^{\mathrm{r}}_y)(P'^{\mathrm{m}}_y - P'^{\mathrm{o}}_y) + (P'^{\prime\prime}_z - P'^{\mathrm{r}}_z)(P'^{\mathrm{m}}_z - P'^{\mathrm{o}}_z) = 0, \quad (5.6)$$

where $\boldsymbol{P}'^{\mathrm{o}} = (P'^{\mathrm{o}}_x, P'^{\mathrm{o}}_y, P'^{\mathrm{o}}_z)$ is the center point of both new 3D eye positions $(\boldsymbol{P}'^{\mathrm{r}}, \boldsymbol{P}'^{\prime\prime})$ calculated as described in Section 5.2, and

$$P'^{\mathrm{m}}_x = \frac{X_m + \Delta X}{F} P'^{\mathrm{m}}_z, \quad P'^{\mathrm{m}}_y = \frac{Y_m + \Delta Y}{F} P'^{\mathrm{m}}_z, \quad (5.7)$$

where $(\Delta X, \Delta Y)$ (Fig. 7(b)) are the deviates from the 2D mouth center positions $(X_{\mathrm{m}}, Y_{\mathrm{m}})$ estimated as described in Section 4. ΔX and ΔY fulfill the following equation:

$$\frac{\Delta Y}{\Delta X} = \frac{Y_{\prime} - Y_{\mathrm{r}}}{X_{\prime} - X_{\mathrm{r}}} =: \tan\alpha, \quad (5.8)$$

with α being the inclination angle of l_{mouth} (Fig. 7(b)).

Eq. (5.7) stands for a perspective projection of a static pinhole camera. Using (5.1) and (5.6)–(5.8) gives the z-coordinate P'^{m}_z of the mouth

$$P'^{\mathrm{m}}_z = F \frac{c_1 c_5 - c_3 c_6}{c_2 c_5 - c_4 c_6} \quad (5.9)$$

and

$$\Delta X = \frac{c_2 c_3 - c_1 c_4}{c_1 c_5 - c_3 c_6}, \quad (5.10)$$

where

$$c_1 := P'^{\mathrm{o}}_x (P'^{\prime\prime}_x - P'^{\mathrm{r}}_x) + P'^{\mathrm{o}}_y (P'^{\prime\prime}_y - P'^{\mathrm{r}}_y) + P'^{\mathrm{o}}_z (P'^{\prime\prime}_z - P'^{\mathrm{r}}_z), \quad (5.11)$$

$$c_2 := X_{\mathrm{m}}(P'^{\prime\prime}_x - P'^{\mathrm{r}}_x) + Y_{\mathrm{m}}(P'^{\prime\prime}_y - P'^{\mathrm{r}}_y) + F(P'^{\prime\prime}_z - P'^{\mathrm{r}}_z), \quad (5.12)$$

$$c_3 := J_{11} P^{\mathrm{m}}_x + J_{12} P^{\mathrm{m}}_y + J_{13} P^{\mathrm{m}}_z, \quad (5.13)$$

$$c_4 := J_{11} X_{\mathrm{m}} + J_{12} Y_{\mathrm{m}} + J_{13} F, \quad (5.14)$$

$$c_5 := J_{11} + J_{12} \tan\alpha. \quad (5.15)$$

$$c_6 := (P'^{\prime\prime}_x - P'^{\mathrm{r}}_x) - (P'^{\prime\prime}_y - P'^{\mathrm{r}}_y) \tan\alpha. \quad (5.16)$$

Substituting (5.10) into (5.8) gives ΔY. Then, the coordinates P'^{m}_x and P'^{m}_y of the mouth are determined by substituting the calculated values of ΔY, ΔX and P'^{m}_z into (5.7).

5.4. Update of the face model's position and size

After the new 3D center positions of the eyes and the mouth of the face model *Candide* have been determined, the face model is scaled and shifted according to Eq. (2.1). Here, the result $\boldsymbol{P}^{*(i)}$ of Eq. (2.1) is the result $\boldsymbol{P}'^{(i)}$ of this step.

6. Estimation of the 3D eye center positions along the view lines and update of the face model's orientation

In Section 5, the face model *Candide* was updated in the *eye–mouth plane* to match the location of the person's face. Furthermore, the *orientation of the face model*, i.e. the norm of the *eye–mouth plane*, is one of the important parameters of the 3D model head shape. Thus, updating the face model's orientation during the sequence will improve the accuracy of face tracking. This problem is addressed in this section.

6.1. Principle of estimation of the face model's orientation

For the simple face model *Candide* as discussed in Section 2, its orientation is determined only by the 3D eye and mouth center positions in the 3D space. However, only using a monocular view of a face is not sufficient to estimate the true positions of the eyes and the mouth of the person in the 3D space. It is known that a parallel movement of the *eye–mouth plane* in the 3D space along the view lines of the eyes and the mouth does not change the face model's orientation. Hence, having those true positions is not a necessary condition to determine the face model's orientation. One of three positions, either eye or mouth, can be fixed during updating the face model's orientation. Here, such a strategy that the mouth center position is fixed and the eye

center positions along their view lines have to be estimated is exploited to update the face model's orientation.

Estimation of the face model's orientation minimizes the mean square luminance differences ΔI in the facial area between the projection of the face model and the real image. The facial area is determined by the projection of the face model onto the image plane, except the eye and mouth areas. These areas are excluded due to the local motion of the eyes and the mouth. Let I_k be the luminance of the frame s_k and $\hat{I}_k(\boldsymbol{P}'', \boldsymbol{P}'^{\mathrm{r}})$ the luminance of the projection [19] of the face model $\hat{s}_k(\boldsymbol{P}'', \boldsymbol{P}'^{\mathrm{r}})$ at time instant k. Here, the positions $P'' = (P''_x, P''_y, P''_z)$, $P'^{\mathrm{r}} = (P''^{\mathrm{r}}_x, P''^{\mathrm{r}}_y, P''^{\mathrm{r}}_z)$ and $P''^{\mathrm{m}} = (P''^{\mathrm{m}}_x, P''^{\mathrm{m}}_y, P''^{\mathrm{m}}_z)$ are the left eye, right eye and mouth center points after the update of the face model's orientation. According to the second assumption in Section 5, the 3D center positions P'', $\boldsymbol{P}'^{\mathrm{r}}$ and $\boldsymbol{P}''^{\mathrm{m}}$ of the eyes and the mouth which are to be estimated build an isosceles triangle (Fig. 6) in the 3D space, i.e.,

$$(\boldsymbol{P}'' - \boldsymbol{P}'^{\mathrm{r}})\,(\boldsymbol{P}''^{\mathrm{m}} - \boldsymbol{P}''^{\mathrm{o}}) = 0. \tag{6.1}$$

Therefore, estimation of the orientation of a person's face can be formulated as follows:

$$\sum (I_k - \hat{I}_k\,(\boldsymbol{P}'', \boldsymbol{P}'^{\mathrm{r}}))^2 \rightarrow \mathrm{MIN}, \tag{6.2}$$

with the three conditions, Eq. (6.1),

$$\boldsymbol{P}''^{\mathrm{m}} = \boldsymbol{P}'^{\mathrm{m}}, \tag{6.3}$$

and

$$\boldsymbol{P}'', \boldsymbol{P}'^{\mathrm{r}} \in \Re^3. \tag{6.4}$$

Condition (6.3) means that the mouth of the face model stays in the same position during updating the face model's orientation.

6.2. Estimation of the 3D eye center positions along the view lines

Estimation of the 3D eye center positions ($\boldsymbol{P}''$, $\boldsymbol{P}'^{\mathrm{r}}$) of the face model minimizes the mean square luminance difference between the projection of the face model $\hat{s}_k(\boldsymbol{P}'', \boldsymbol{P}'^{\mathrm{r}})$ and the real image s_k. It is assumed that objects have diffusely reflecting surfaces. Furthermore, diffuse illumination of the scene is assumed. Hence, color parameters are temporarily constant. With an observation point [19] $\boldsymbol{O}'^{(j)}_k = (\boldsymbol{P}'^{(j)}_k, \boldsymbol{g}^{(j)}, \boldsymbol{I}^{(j)})$ at time instant k projected onto the image plane at $\boldsymbol{p}'^{(j)}_k$ and the same observation point after the update of the face model's orientation $\boldsymbol{O}''^{(j)}_k = (\boldsymbol{P}''^{(j)}_k, \boldsymbol{g}^{(j)}, \boldsymbol{I}^{(j)})$ projected onto $\boldsymbol{p}''^{(j)}_k$, the luminance difference $\Delta I^{(j)}$ between the projection of the face model $\hat{s}_k$ and the real image s_k at position $\boldsymbol{p}'^{(j)}_k$ is then related to both eye coordinates by

$$\Delta I^{(j)} = s_k(\boldsymbol{p}'^{(j)}_k) - \hat{s}_k(\boldsymbol{p}'^{(j)}_k) = (g^{(j)}_x, g^{(j)}_y)^{\mathrm{T}}\,(\boldsymbol{p}''^{(j)}_k - \boldsymbol{p}'^{(j)}_k). \tag{6.5}$$

Substituting image coordinates by model world coordinates with a perspective projection of a pinhole camera yields

$$\Delta I^{(j)} = F\,g^{(j)}_x \left(\frac{P''^{(j)}_{x,k}}{P''^{(j)}_{z,k}} - \frac{P'^{(j)}_{x,k}}{P'^{(j)}_{z,k}}\right) + F\,g^{(j)}_y \left(\frac{P''^{(j)}_{y,k}}{P''^{(j)}_{z,k}} - \frac{P'^{(j)}_{y,k}}{P'^{(j)}_{z,k}}\right). \tag{6.6}$$

The position $\boldsymbol{P}'^{(j)}_k$ of the observation point $O'^{(j)}_k$ is known. Relating $\boldsymbol{P}'^{(j)}_k$ to $\boldsymbol{P}''^{(j)}_k$ by means of Eq. (2.1), a non-linear equation with the known parameters $\Delta I^{(j)}$, $\boldsymbol{g}^{(j)}$ and F and two unknown z-coordinates of both eyes (P'_z, P''^{r}_z) results. Let (ΔP^{r}_z, $\Delta P'_z$) be the alterations of z-coordinates of both eyes. For sufficiently small alterations (ΔP^{r}_z, $\Delta P'_z$), the linearized equation using the Taylor series expansion is

$$\Delta I^{(j)} = \left.\frac{\partial \Delta I^{(j)}}{\partial P''_z}\right|_{\substack{P''_z = P'_z \\ P''^{\mathrm{r}}_z = P'^{\mathrm{r}}_z}} \Delta P'_z + \left.\frac{\partial \Delta I^{(j)}}{\partial P''^{\mathrm{r}}_z}\right|_{\substack{P''_z = P'_z \\ P''^{\mathrm{r}}_z = P'^{\mathrm{r}}_z}} \Delta P^{\mathrm{r}}_z, \tag{6.7}$$

where

$$\left.\frac{\partial \Delta I^{(j)}}{\partial P''_z}\right|_{\substack{P''_z = P'_z \\ P''^{\mathrm{r}}_z = P'^{\mathrm{r}}_z}} \quad \text{and} \quad \left.\frac{\partial \Delta I^{(j)}}{\partial P''^{\mathrm{r}}_z}\right|_{\substack{P''_z = P'_z \\ P''^{\mathrm{r}}_z = P'^{\mathrm{r}}_z}}$$

depend on the observation point $\boldsymbol{O}'^{(j)}_k$ and the eye and mouth positions ($\boldsymbol{P}'$, $\boldsymbol{P}'^{\mathrm{r}}$, $\boldsymbol{P}'^{\mathrm{m}}$).

Furthermore, from Eq. (6.1), there is

$$\left.\frac{\partial P''^{\mathrm{r}}_z}{\partial P''_z}\right|_{\substack{P''_z = P'_z \\ P''^{\mathrm{r}}_z = P'^{\mathrm{r}}_z}} =: R(\boldsymbol{P}', \boldsymbol{P}'^{\mathrm{r}}, \boldsymbol{P}'^{\mathrm{m}}), \tag{6.8}$$

where $R(\boldsymbol{P}'^{l}, \boldsymbol{P}'^{r}, \boldsymbol{P}'^{m})$ is a coefficient only depending on the eye and mouth center positions ($\boldsymbol{P}'^{l}$, $\boldsymbol{P}'^{r}$, $\boldsymbol{P}'^{m}$). For sufficiently small alterations (ΔP_z^r, ΔP_z^l), the alteration of z-coordinates of the right eye is equal to

$$\Delta P_z^r = R(\boldsymbol{P}'^{l}, \boldsymbol{P}'^{r}, \boldsymbol{P}'^{m})\, \Delta P_z^l. \tag{6.9}$$

Due to the geometric condition (6.1), both eye center positions depend on each other. According to the results of Appendix A, substituting Eq. (6.9) into Eq. (6.7), we have

$\theta_l > \theta_r$:

$$\Delta I^{(j)} = \left[\left. \frac{\partial \Delta I^{(j)}}{\partial P_z''^{l}} \right|_{\substack{P_z''^{l} = P_z'^{l} \\ P_z''^{r} = P_z'^{r}}} + \left. \frac{\partial \Delta I^{(j)}}{\partial P_z''^{r}} \right|_{\substack{P_z''^{l} = P_z'^{l} \\ P_z''^{r} = P_z'^{r}}} \cdot R(\boldsymbol{P}'^{l}, \boldsymbol{P}'^{r}, \boldsymbol{P}'^{m}) \right] \Delta P_z^l , \tag{6.10}$$

$\theta_l < \theta^r$ *or* $\theta_l = \theta_r$:

$$\Delta I^{(j)} = \left[\frac{1}{R(\boldsymbol{P}'^{l}, \boldsymbol{P}'^{r}, \boldsymbol{P}'^{m})} \left. \frac{\partial \Delta I^{(j)}}{\partial P_z''^{l}} \right|_{\substack{P_z''^{l} = P_z'^{l} \\ P_z''^{r} = P_z'^{r}}} + \left. \frac{\partial \Delta I^{(j)}}{\partial P_z''^{r}} \right|_{\substack{P_z''^{l} = P_z'^{l} \\ P_z''^{r} = P_z'^{r}}} \right] \Delta P_z^r, \tag{6.11}$$

where the two symbols θ_l and θ_r are defined in Fig. 8 (see Appendix A).

In order to get the reliable estimate, Eq. (6.10) or Eq. (6.11) has to be established for several hundred observation points. The residuum of the equation system in the different cases is then minimized by linear regression:

$$\sum_{\boldsymbol{O}'(j)} (\Delta I^{(j)})^2 \rightarrow \text{MIN}. \tag{6.12}$$

In order to make the estimation more robust, only observation points should be used for which the following inequation is satisfied [19]:

$$|\Delta I^{(j)}| < \sigma_I, \tag{6.13}$$

where σ_I is the standard deviation of all residuals $\Delta I^{(j)}$ according to Eq. (6.10) or Eq. (6.11).

After that, the z-coordinate of one eye of the face model is compensated according to the different cases, i.e.,

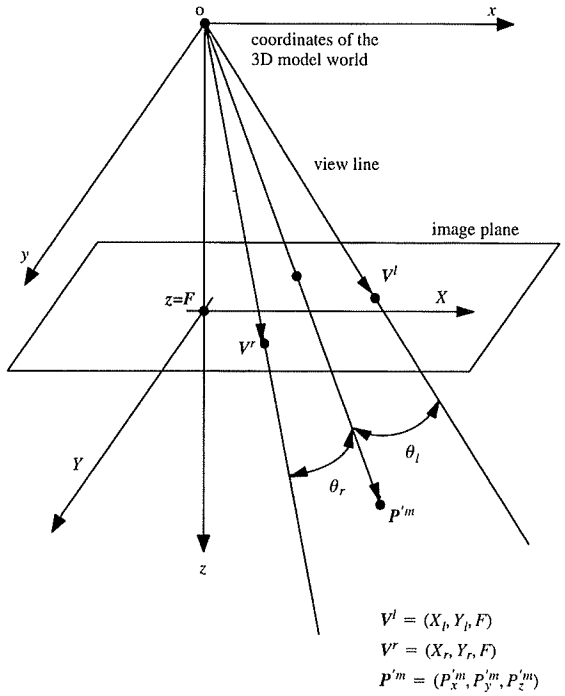


Fig. 8. Definition of the angles θ_l and θ_r.

$$\theta_l > \theta_r: \quad P_z''^{l} = P_z'^{l} - \Delta P_z^l, \tag{6.14}$$

$$\theta_l < \theta^r \text{ or } \theta_l = \theta_r: \quad P_z''^{r} = P_z'^{r} - \Delta P_z^r, \tag{6.15}$$

whereas the z-coordinate of the other eye is calculated based on Eq. (6.1). In general, Eq. (6.1) gives two values of the eye z-coordinate. That value with the smaller mean square luminance difference between the projection of the face model and the real image, except the eye and mouth areas, is selected as the eye z-coordinate. Then, the x- and y-coordinate of both eyes are calculated according to a perspective projection of a pinhole camera (Eqs. (5.2) and (5.3)), and the face model is moved according to Eq. (2.1). Afterwards, a new set of estimation equations is established, which give the new alterations of the eye z-coordinates. Due to the linearization, the 3D eye coordinates have to be estimated iteratively.

6.3. Update of the face model's orientation

After the 3D eye center positions of the face model *Candide* along the view lines have been estimated, the face model's orientation is updated according to Eq. (2.1). Here, the result $\boldsymbol{P}^{*(i)}$ of Eq. (2.1) is the result $\boldsymbol{P}''^{(i)}$ of this step.

7. Experimental results

The proposed algorithm for face tracking has been tested with natural image sequences and synthetic images. The natural image sequences are *Miss America* and *Akiyo* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz. For generating a synthetic image, a wireframe (Fig. 9(e)) including the face model *Candide* is used, which was calculated with the coder in [9] and adapted on an original frame of the sequence *Miss America* (Fig. 9(a)). After projection of the texture onto the wireframe, it is moved in the 3D space and a synthetic image is created by an image synthesis. In the experiments with the synthetic images, the motion back to the original image has to be estimated.

7.1. Results with synthetic images

The proposed algorithm has been tested with synthetic images. The improvements of the proposed algorithm are shown with the accuracy of the estimated face model's orientation, which is represented by the angles $(\gamma_x, \gamma_y, \gamma_z)$ between the norm of the *eye–mouth plane* and the x-, y- and z-axis of the coordinate system, respectively.

7.1.1. Simulation for the update of the face model's orientation

First, only a part of the proposed algorithm, the update of the face model's orientation, is tested. The synthetic image is generated by an image synthesis after moving both eyes of the textured face model *Candide* along their view lines. The mean square difference of the luminance between the synthetic image and the original image is 4.584122, which originates from the change of the face model's orientation. Table 1 shows the simulation results. Compared to global head motion compensation [9], the absolute error between the face model's orientation in the original image and the face model's orientation in the synthesized image is reduced from (1.937, 3.284, 2.147) to (0.173, 0.060, 0.128) by the proposed algorithm for the update of the face model's orientation. The mean square error (MSE) of the luminance between the original image and the synthesized image after global head motion compensation [9] is 1.726257, while the MSE of the luminance between the original image and the synthesized image after the proposed algorithm for the update of the face model's orientation is 0.008296.

7.1.2. Simulation for the whole face tracking system

Here, the complete algorithm for face tracking is tested. The synthetic image (Fig. 9(b)) is generated by an image synthesis after rotation and translation of the textured wireframe (Fig. 9(e)) in the 3D space and additional motion of the face model *Candide* along the eye view lines. The mean square difference of the luminance between the synthetic image and the original image is 52.035807. Table 2 shows the simulation results. Compared to face tracking by global head motion compensation only [9], the absolute error between the face model's orientation in the original image and the face model's orientation in the synthesized image is reduced from (2.185, 1.323, 0.201) to (0.063, 0.338, 0.232) by the proposed algorithm for face tracking. The MSE of the luminance between the original image and the synthesized image (Fig. 9(c)) after face tracking by global head motion compensation only [9] is 0.630228, while the MSE of the luminance between the original image and the synthesized image (Fig. 9(d)) after face tracking by the proposed algorithm is 0.553297.

7.2. Results with natural videophone sequences

The proposed algorithm has been combined with the image analysis of a KBASC according to [9] and tested with different image sequences. Here, results for the two sequences *Miss America* and *Akiyo* using 50 frames for each are given. The goal of the face tracking is that the projection of the face

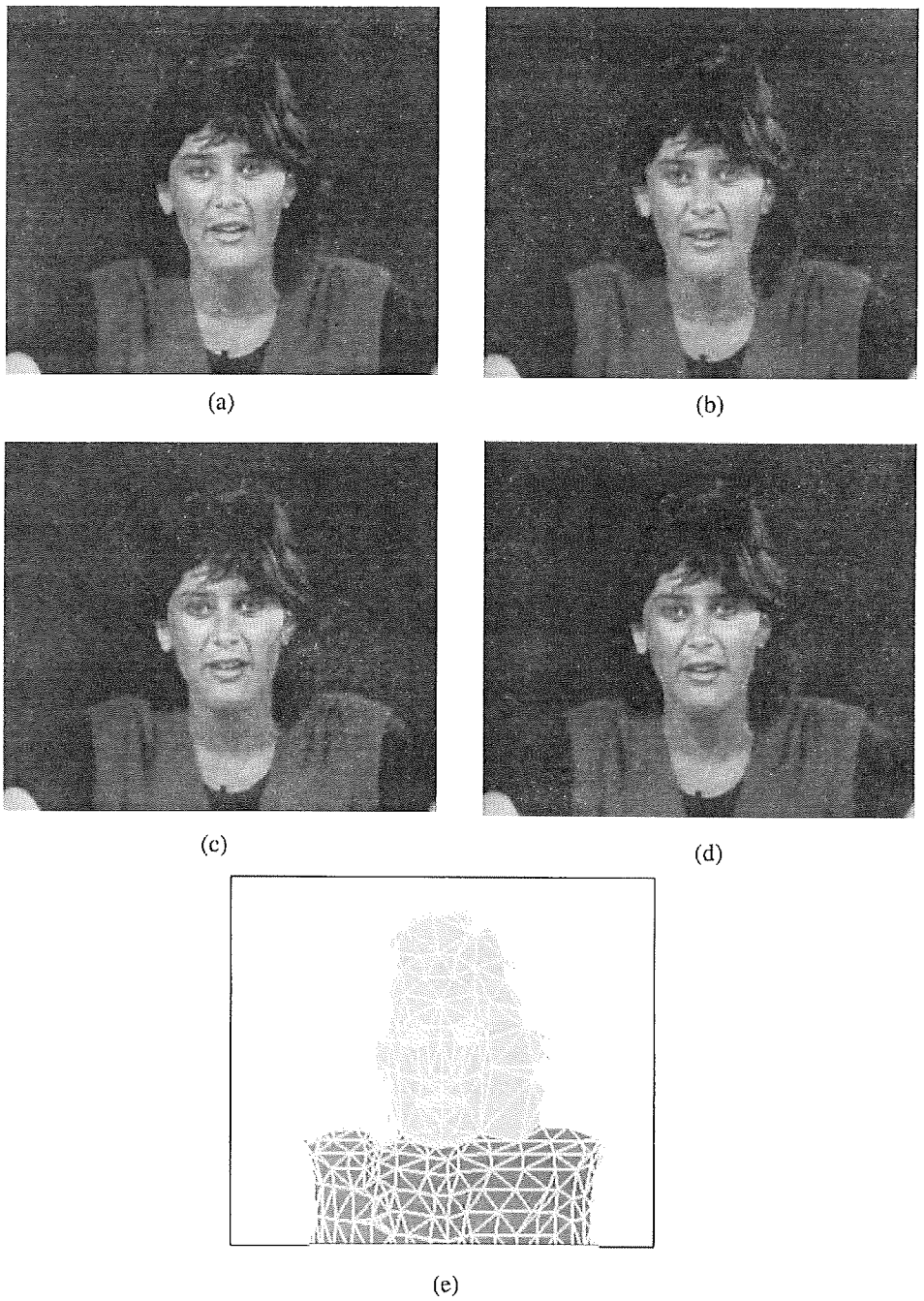


Fig. 9. Simulation results for tracking a face: (a) original image; (b) synthetic image; (c) synthesized image after face tracking by global head motion compensation only [9]; (d) synthesized image after face tracking by the proposed algorithm; (e) a wireframe with components *Candide*, *head* and *shoulders*.

Table 1
Simulation results for the update of the face model's orientation

	Orientation of the face model *Candide* (degree)			
	In the original image	In the synthetic image	After global head motion compensation [9] (absolute error)	After update of the face model's orientation (absolute error)
γ_x	91.915	89.744	89.978 (1.937)	92.088 (0.173)
γ_y	91.044	95.052	94.328 (3.284)	90.984 (0.060)
γ_z	2.181	5.058	4.328 (2.147)	2.309 (0.128)

Table 2
Simulation results for the whole face tracking system

	Orientation of the face model *Candide* (degree)			
	In the original image	In the synthetic image	After face tracking by global head motion compensation [9] (absolute error)	After face tracking by the proposed algorithm (absolute error)
γ_x	91.915	88.266	89.730 (2.185)	91.978 (0.063)
γ_y	91.044	93.708	92.367 (1.323)	91.382 (0.338)
γ_z	2.181	4.094	2.384 (0.201)	2.413 (0.232)

model *Candide* onto the image plane matches the location of the real face in the scene. Because the luminance differences in the regions of the eyes and the mouth are not taken into consideration during updating the face model's orientation, it may be that the MSE of the luminance after the update of the face model's shape might be larger than that after global head motion compensation only, although the face tracking accuracy is improved. Hence, the criterion MSE is not suitable to measure the improvement of the face tracking algorithm. In order to evaluate the improvement with respect to the projection of the face model *Candide* onto the image plane, the maximum position errors for the eyes and the mouth at image frame k and their average position errors,

$$f^{\max}_{\text{eye},k} = \max\left[|X_{\ell,k} - X^{t}_{\ell,k}|, |Y_{\ell,k} - Y^{t}_{\ell,k}|, |X_{r,k} - X^{t}_{r,k}|, |Y_{r} - Y^{t}_{r,k}|\right], \quad (7.1)$$

$$f^{\max}_{\text{mouth},k} = \max\left[|X_{m,k} - X^{t}_{m,k}|, |Y_{m,k} - Y^{t}_{m,k}|\right], \quad (7.2)$$

$$\bar{f}_{\text{eye}} = \frac{1}{N_{\text{eye}}} \sum_{k=1}^{N_{\text{eye}}} f^{\max}_{\text{eye},k}, \quad (7.3)$$

$$\bar{f}_{\text{mouth}} = \frac{1}{N} \sum_{k=1}^{N} f^{\max}_{\text{mouth},k}, \quad (7.4)$$

are introduced, where $(X^{t}_{\ell,k}, Y^{t}_{\ell,k})$, $(X^{t}_{r,k}, Y^{t}_{r,k})$ and $(X^{t}_{m,k}, Y^{t}_{m,k})$ are the true image coordinates of the open eyes and the mouth at image frame k (manually determined), $(X_{\ell,k}, Y_{\ell,k})$, $(X_{r,k}, Y_{r,k})$ and $(X_{m,k}, Y_{m,k})$ are the coordinates of the eyes and the mouth of the face model projected onto the image plane at image frame k. N is the number of images. N_{eye} is the number of images in which the eyes are open.

For the test sequence *Miss America*, the face model *Candide* is adapted after the second frame. The motion of this sequence is mainly parallel to the image plane. Figs. 10 and 11 show the maximum position errors $f^{\max}_{\text{eye},k}$ and $f^{\max}_{\text{mouth},k}$ over the frame k, respectively. In Fig. 10, the missing data originate from the fact that the eyes in these frames

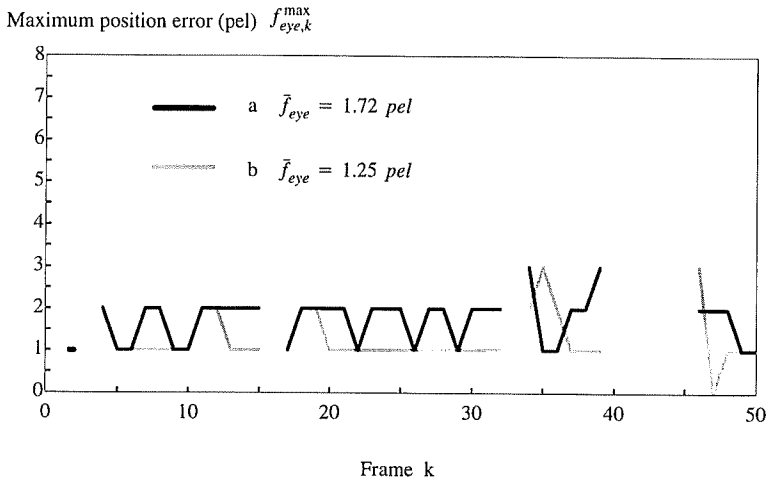


Fig. 10. Maximum position error of both eyes for the test sequence *Miss America* (CIF, 10 Hz), the face model is adapted after the second frame. (a) Face tracking by global head motion compensation only [9]. The average eye position error is 1.72 pel. (b) Face tracking by the proposed algorithm. The average eye position error is 1.25 pel.

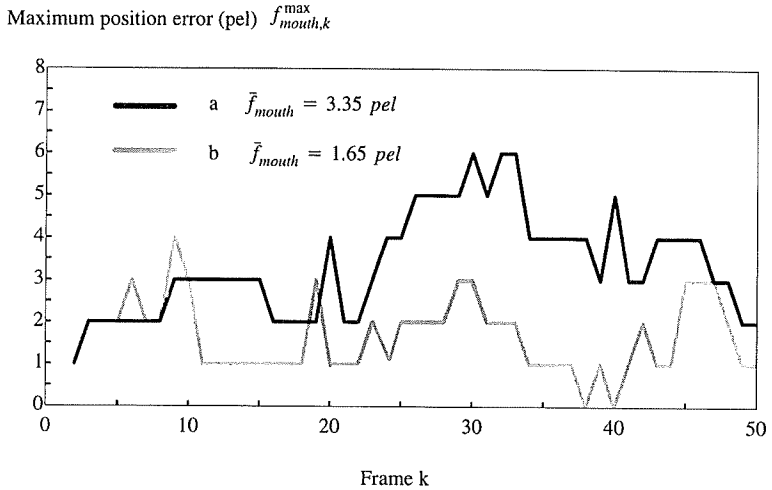


Fig. 11. Maximum position error of the mouth for the test sequence *Miss America* (CIF, 10 Hz), the face model is adapted after the second frame. (a) Face tracking by global head motion compensation only [9]. The average mouth position error is 3.35 pel. (b) Face tracking by the proposed algorithm. The average mouth position error is 1.65 pel.

are closed and the true center coordinates cannot be obtained. Compared to face tracking by global head motion compensation only [9], the average position error for the eyes $\bar{f}_{\text{eye}}$ is reduced from 1.72 to 1.35 pel by updating the face model's position and size according to the steps described in Sections 4 and 5, and this error value is further reduced to 1.25 pel by additionally updating the face model's orientation according to the step described in Section 6. Similarly, compared to face tracking

by global head motion compensation only [9], the average position error for the mouth $\bar{f}_{mouth}$ is reduced from 3.35 to 2.02 pel by the steps of Sections 4 and 5 and further reduced to 1.65 pel by the step of Section 6. Fig. 12 shows the eye and mouth center positions of the face model projected onto the image plane for the frames 38–49 with face tracking by global head motion compensation only [9] and by the proposed algorithm.

(a) (b)

Fig. 12. Eye and mouth center positions for the frames 38–49 of the test sequence *Miss America* (CIF, 10 Hz) with face tracking by (a) global head motion compensation only [9], (b) the proposed algorithm.

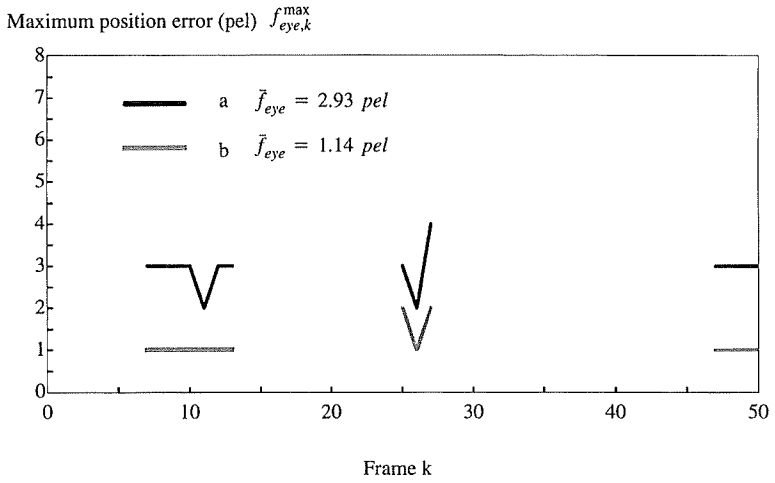


Fig. 13. Maximum position error of both eyes for the test sequence *Akiyo* (CIF, 10 Hz), the face model is adapted after the seventh frame. (a) Face tracking by global head motion compensation only [9]. The average eye position error is 2.93 pel. (b) Face tracking by the proposed algorithm. The average eye position error is 1.14 pel.

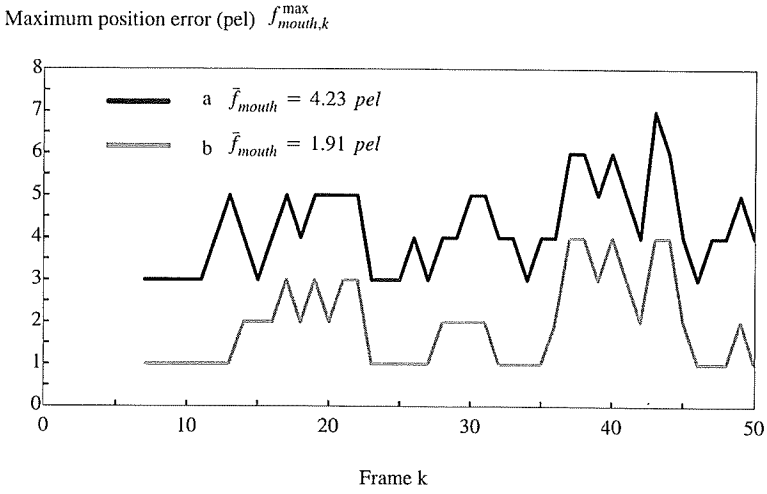


Fig. 14. Maximum position error of the mouth for the test sequence *Akiyo* (CIF, 10 Hz), the face model is adapted after the seventh frame. (a) Face tracking by global head motion compensation only [9]. The average mouth position error is 4.23 pel. (b) Face tracking by the proposed algorithm. The average mouth position error is 1.91 pel.

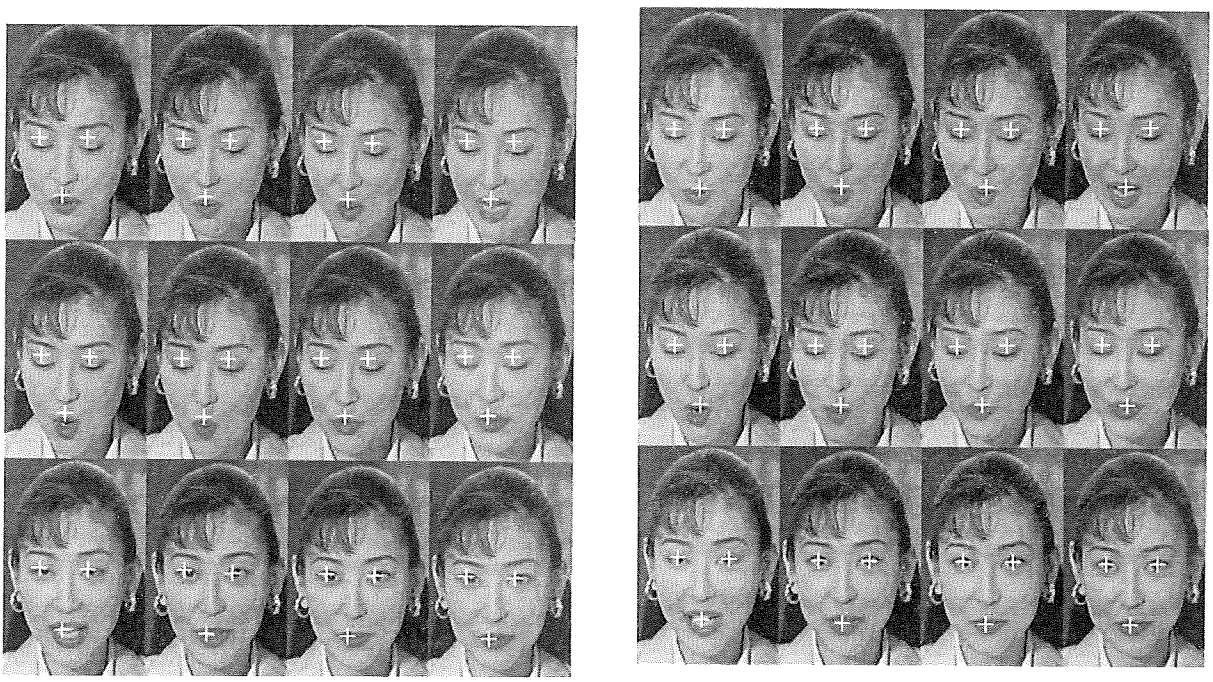


Fig. 15. Eye and mouth center positions for the frames 39–50 of the test sequence *Akiyo* (CIF, 10 Hz) with face tracking by (a) global head motion compensation only [9], (b) the proposed algorithm.

For the test sequence *Akiyo*, the face model *Candide* is adapted after the seventh frame. In this sequence, the person often looks towards the bottom and the eyes are closed in many images. Figs. 13 and 14 show the maximum position errors $f^{\max}_{\text{eye},k}$ and $f^{\max}_{\text{mouth},k}$ over the frame k, respectively. In Fig. 13, the missing data originate from the fact that the eyes in these frames are closed and the true center coordinates cannot be obtained. Compared to face tracking by global head motion compensation only [9], the average position error for the eyes $\bar{f}_{\text{eye}}$ is reduced from 2.93 to 1.14 pel by the proposed face tracking algorithm and the average position error for the mouth $\bar{f}_{\text{mouth}}$ is reduced from 4.23 to 1.91 pel. Fig. 15 shows the eye and mouth center positions of the face model projected onto the image plane for the frames 39–50 with face tracking by global head motion compensation only [9] and by the proposed algorithm.

7.3. *Increase of the computation complexity*

In order to measure the additional complexity introduced by the proposed algorithm, the computation time is evaluated. Experimental results applying the test sequence *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz show that the computation time for the image analysis increases by about 10% compared to the image analysis scheme used in [9].

8. Conclusions

In this paper, a new algorithm for face tracking in a knowledge-based coder for videophone sequences has been presented. It combines global head motion compensation and update of the face model's shape. Global head motion compensation gives a coarse face tracking, while the shape update is used to further improve the face tracking's accuracy. As a first stage of this algorithm, the method of tracking a face by global head motion compensation from [9] is used. Then, template matching and feature point extraction techniques are used to estimate the 2D eye and mouth center positions of the person's face in the image plane. By means of these estimated 2D center positions, the shape of the face model is updated during the sequence. For shape update, not only the locations of the eyes and the mouth of the face model are adapted to match the positions of the real eye and mouth in the sequence but also the orientation of the face model is updated.

In order to evaluate the face tracking accuracy for knowledge-based coding of videophone sequences, the developed algorithm was combined with the image analysis of a knowledge-based coding scheme according to [9]. Typical head-and-shoulder videophone sequences with a spatial resolution according to CIF and a frame rate of 10 Hz have been investigated. For evaluation, error criteria have been introduced which give the position errors of the eyes and the mouth averaged over a whole sequence. Compared to face tracking by global head motion compensation only [9], the proposed face tracking algorithm reduces the average position errors for the eyes and the mouth by 48% and 53%, respectively, for the used test sequences. These experimental results show that the proposed algorithm allows a more accurate tracking than by global head motion compensation only [9]. Thus, a more precise modelling of 3D objects is possible which is required for combined coding of natural and synthetic scene contents in the framework of MPEG-4 Synthetic/Natural Hybrid Coding (SNHC).

With the accurate tracking of the face, it is easier to estimate the other facial feature points or contours, e.g. the positions of the eyelids and the lips of the mouth. Furthermore, how the coding efficiency of KBASC is affected by the more accurate tracking of the face is another important issue. These topics will be investigated further.

Acknowledgements

The author wishes to thank Prof. Dr.-Ing. H.G. Musmann for encouraging this work and many helpful discussions on KBASC. Furthermore, the author thanks Dipl.-Ing. M. Kampmann for his fruitful discussions on image analysis in KBASC. This research was supported by the Deutscher Akademischer Austauschdienst (DAAD), Germany.

Appendix A

Eq. (6.1) in Section 6 is rewritten as follows:

$$[X_l^2 + Y_l^2 + F^2]\, P_z''^2 - 2F\,(P_x'^{\mathrm{m}} X_l + P_y'^{\mathrm{m}} Y_l + P_z'^{\mathrm{m}} F)\, P_z'' = [X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2]\, P_z''^2 - 2F\,(P_x'^{\mathrm{m}} X_{\mathrm{r}} + P_y'^{\mathrm{m}} Y_{\mathrm{r}} + P_z'^{\mathrm{m}} F)\, P_z''^{\mathrm{r}}. \quad (A.1)$$

It is known that the coordinates $(P_z'', P_z''^{\mathrm{r}})$ must be real. *If P_z'' is given*, the coefficients of Eq. (A.1) must satisfy inequality (A.2) in order to have a real value $P_z''^{\mathrm{r}}$ from Eq. (A.1):

$$(X_l^2 + Y_l^2 + F^2)(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2)\, P_z''^2 - 2F(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2) \times (P_x'^{\mathrm{m}} X_l + P_y'^{\mathrm{m}} Y_l + P_z'^{\mathrm{m}} F)\, P_z'' + F^2(P_x'^{\mathrm{m}} X_r + P_y'^{\mathrm{m}} Y_r + P_z'^{\mathrm{m}} F)^2 \geqslant 0, \quad (A.2)$$

i.e. P_z'' must satisfy inequality (A.2) so that we have a real value $P_z''^{\mathrm{r}}$. Now, the question is how we can ensure to have a real value P_z'' under the condition of inequality (A.2). In order to answer this question, we consider the equality condition of (A.2), i.e.,

$$(X_l^2 + Y_l^2 + F^2)(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2)\, P_z''^2 - 2F(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2) \times (P_x'^{\mathrm{m}} X_l + P_y'^{\mathrm{m}} Y_l + P_z'^{\mathrm{m}} F)\, P_z'' + F^2\,(P_x'^{\mathrm{m}} X_r + P_y'^{\mathrm{m}} Y_{\mathrm{r}} + P_z'^{\mathrm{m}} F)^2 = 0. \quad (A.3)$$

Solving Eq. (A.3) gives two values $P_z''^1$ and $P_z''^2$ $(P_z''^1 \leqslant P_z''^2)$. Because inequality (A.2) is a concave function, inequality (A.2) holds if $P_z'' \leqslant P_z''^1$ or $P_z'' \geqslant P_z''^2$, i.e. any values in the open interval $(P_z''^1, P_z''^2)$ cannot be selected as P_z''. In this case, both eye positions of the face model along their view lines can only be located on the dark dashed lines (Fig. 16). In order to have real values $P_z''^1$ and $P_z''^2$, the coefficients of Eq. (A.3) must satisfy the following condition:

$$F^2(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2)^2\,(P_x'^{\mathrm{m}} X_l + P_y'^{\mathrm{m}} Y_l + P_z'^{\mathrm{m}} F)^2 - (X_l^2 + Y_l^2 + F^2)(X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2)F^2 \times (P_x'^{\mathrm{m}} X_{\mathrm{r}} + P_y'^{\mathrm{m}} Y_{\mathrm{r}} + P_z'^{\mathrm{m}} F)^2 \geqslant 0. \quad (A.4)$$

If the equality condition of (A.4) is satisfied, $P_z''^1 = P_z''^2$, i.e. any real value can be selected as P_z''. In this case, both eye positions of the face model along their view lines can be located on the dark dashed lines (Fig. 17), i.e. along the complete view lines. Inequality (A.4) can be rewritten as follows:

$$\left|\frac{P_x'^{\mathrm{m}} X_l + P_y'^{\mathrm{m}} Y_l + P_z'^{\mathrm{m}} F}{\sqrt{X_l^2 + Y_l^2 + F^2}\,\sqrt{P_x'^{\mathrm{m}^2} + P_y'^{\mathrm{m}^2} + P_z'^{\mathrm{m}^2}}}\right| \geqslant \left|\frac{P_x'^{\mathrm{m}} X_{\mathrm{r}} + P_y'^{\mathrm{m}} Y_{\mathrm{r}} + P_z'^{\mathrm{m}} F}{\sqrt{X_{\mathrm{r}}^2 + Y_{\mathrm{r}}^2 + F^2}\,\sqrt{P_x'^{\mathrm{m}^2} + P_y'^{\mathrm{m}^2} + P_z'^{\mathrm{m}^2}}}\right|. \quad (A.5)$$

In order to simplify expression (A.5), let vectors $\boldsymbol{V}^l$, $\boldsymbol{V}^{\mathrm{r}}$ and $\boldsymbol{P}'^{\mathrm{m}}$ be the points (X_l, Y_l, F), $(X_{\mathrm{r}}, Y_{\mathrm{r}}, F)$

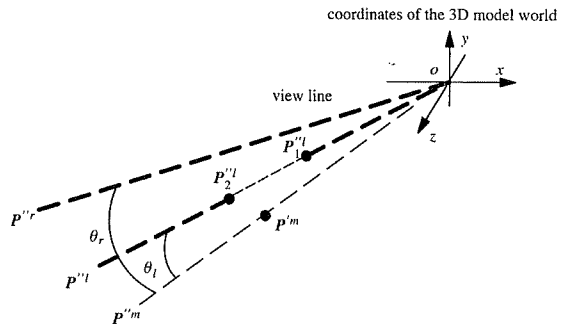


Fig. 16. Possible center positions of the eyes located on the dark dashed lines for $\theta_l < \theta_{\mathrm{r}}$. OP''^{r}: view line of the right eye; OP''^{l}: view line of the left eye; OP''^{m}: view line of the mouth.

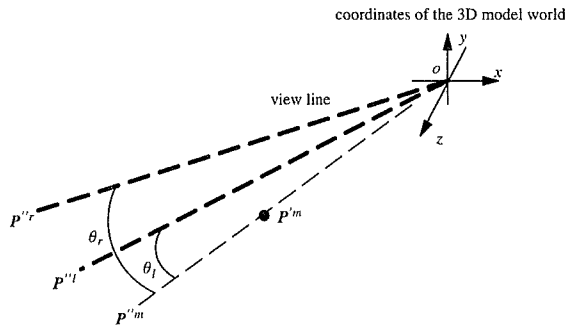


Fig. 17. Possible center positions of the eyes located on the dark dashed lines for $\theta_l = \theta_{\mathrm{r}}$. OP''^{r}: view line of the right eye; OP''^{l}: view line of the left eye; OP''^{m}: view line of the mouth.

and $(P_x'^m, P_y'^m, P_z'^m)$ in the 3D coordinate system (Fig. 8), respectively. With these vectors, we have

$$\frac{|\boldsymbol{P}'^m \boldsymbol{V}'|}{\|\boldsymbol{P}'^m\| \|\boldsymbol{V}'\|} \geqslant \frac{|\boldsymbol{P}'^m \boldsymbol{V}^r|}{\|\boldsymbol{P}'^m\| \|\boldsymbol{V}^r\|}. \tag{A.6}$$

It is known that the inner product AB of two vectors A and B is defined as follows:

$$\boldsymbol{AB} = \|\boldsymbol{A}\| \|\boldsymbol{B}\| \cos\theta, \tag{A.7}$$

where θ $(0 \leqslant \theta \leqslant \pi)$ is the angle surrounded by two vectors A and B. Substituting (A.7) into inequality (A.6) yields

$$|\cos\theta_l| \geqslant |\cos\theta_r|, \tag{A.8}$$

where θ_l is the angle between the view line of the left eye and the view line of the mouth and θ_r the angle between the view line of the right eye and the view line of the mouth (Fig. 8). Because cosine is a mono-decreasing function in the interval $[0, \pi/2,]$ we have

$$\theta_l \leqslant \theta_r \tag{A.9}$$

from inequality (A.8) with θ_l $(0 \leqslant \theta_l \leqslant \pi/2)$ and θ_r $(0 \leqslant \theta_r \leqslant \pi/2)$.

In the case where $P_z''^r$ is given, it yields

$$\theta_l > \theta_r. \tag{A.10}$$

In this case, both eye positions of the face model along their view lines can only be located on the dark dashed lines (Fig. 18).

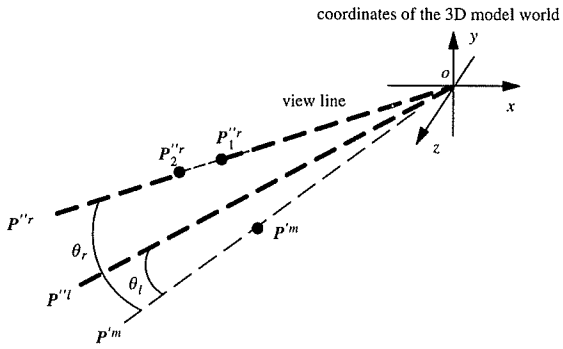


Fig. 18. Possible center positions of the eyes located on the dark dashed lines for $\theta_l > \theta_r$. OP''^r: view line of the right eye; OP''^l: view line of the left eye; OP''^m: view line of the mouth.

References

[1] J.K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images – A review", *Proc. IEEE*, Vol. 76, No. 8, August 1988, pp. 917–935.

[2] K. Aizawa and T.S. Huang, "Model-based image coding: advanced video coding techniques for very low bit-rate applications", *Proc. IEEE*, Vol. 83, No. 2, February 1995, pp. 259–271.

[3] G. Bozdagi, A.M. Tekalp and L. Onural, "3D motion estimation and wireframe adaptation including photometric effects for model-based coding of facial image sequences", *IEEE Trans. Circuits Systems Video Technol.*, Vol. 4, No. 3, June 1994, pp. 246–256.

[4] G. Bozdagi, A.M. Tekalp and L. Onural, "An improvement to MBASIC algorithm for 3-D motion and depth estimation", *IEEE Trans. Image Process.*, Vol. 3, No. 5, September 1994, pp. 711–716.

[5] 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 Systems Video Technol.*, Vol. 4, No. 3, June 1994, pp. 257–275.

[6] T.S. Huang and A.N. Netravali, "Motion and structure from feature correspondences: A review", *Proc. IEEE*, Vol. 82, No. 2, February 1994, pp. 252–268.

[7] ISO/IEC, MPEG-4 synthetic/natural hybrid coding: Call for proposals, ISO/IEC JTC1/SC29/WG11 N1195, March 1996.

[8] ISO/IEC, MPEG-4 synthetic/natural hybrid coding: Proposal package description, ISO/IEC JTC1/SC29/WG11 N1199, March 1996.

[9] 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, March 1997, pp. 201–220.

[10] F. Kappei and G. Heipel, "3D model based image coding", *Picture Coding Symp. (PCS '88)*, Torino, Italy, Paper 4.2, September 1988.

[11] F. Kappei and C.-E. Liedtke, "Modelling of a natural 3-D scene consisting of moving objects from a sequence of monocular TV images", *Real Time Image Processing: Concepts and Technologies*, Cannes, France, SPIE-Vol. 860, November 1987, pp. 126–132.

[12] R. Koch, "Adaptation of a 3D facial mask to human faces in videophone sequences using model based image analysis", *Picture Coding Symp. (PCS '91)*, Tokyo, Japan, September 1991, pp. 285–288.

[13] H. Li, Low bitrate image sequence coding, Ph.D. Dissertation, Department of Electrical Engineering, Linköping University, Sweden, 1993, pp. 139–152.

[14] H. Li, R. Forchheimer, "Two-view facial movement estimation", *IEEE Trans. Circuits Systems Video Technol.*, Vol. 4, No. 3, June 1994, pp. 276–287.

[15] H. Li, A. Lundmark and R. Forchheimer, "Image sequence coding at very low bitrates: A review", *IEEE Trans. Image Process.*, Vol. 3, No. 5, September 1994, pp. 589–609.

[16] G. Martínez, "Shape estimation of articulated 3D objects for object-based analysis-synthesis coding (OBASC)", *Signal Processing: Image Communication*, Vol. 9, No. 3, March 1997, pp. 175–199.

[17] H.G. Musmann, "A layered coding system for very low bit rate video coding", *Signal Processing: Image Communication*, Vol. 7, Nos. 4–6, November 1995, pp. 267–278.

[18] K. Ohmura, A. Tomono and Y. Kobayashi, "Method of detecting face direction using image processing for human interface", *Visual Communications and Image Processing'88*, Cambridge, MA, SPIE Vol. 1001, November 1988, pp. 625–632.

[19] 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, May 1994, pp. 143–161.

[20] J. Ostermann, "Object-based analysis-synthesis coding (OBASC) based on the source model of moving flexible 3D objects", *IEEE Trans. Image Process.*, Vol. 3, No. 5, September 1994, pp. 705–711.

[21] J. Ostermann and M. Kampmann, "Automatic adaptation of a face model in an analysis-synthesis coder based on moving 3D-objects", *Proc. Internat. Workshop on Coding Techniques for Very Low Bit-rate Video (VLBV 94)*, Colchester, United Kingdom, Paper 2.4, April 1994.

[22] D.E. Pearson, "Developments in model-based video coding", *Proc. IEEE*, Vol. 83, No. 6, June 1995, pp. 892–906.

[23] 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", *Visual Communications and Image Processing'93*, Cambridge, MA, SPIE Vol 2904, November 1993, pp. 1516–1527.

[24] R. Rydfalk, CANDIDE, A parameterised face, Internal Report Lith-ISY-I-0866, Linköping University, Linköping, Sweden, 1987.

[25] J. Stauder, "Global-shape from motion and shading for 3D-object-based analysis-synthesis coding", *Proc. Internat. Workshop on Coding Techniques for Very Low Bit-Rate Video (VLBV '95)*, Tokyo, Japan, paper H-3, November 1995.

[26] L. Zhang, "Tracking a face in a knowledge-based analysis-synthesis coder", *Proc. Internat. Workshop on Coding Techniques for Very Low Bit-Rate Video (VLBV '95)*, Tokyo, Japan, Paper A-6, November 1995.

Liang Zhang was born in Zhejiang, V.R.China on 7 November 1961. He received the M.Sc. degree in Electrical Engineering in 1986 from Shanghai Jiaotong University. He was an assistant from 1987 to 1988 and a lecturer from 1989 to 1992 in Electrical Engineering at Shanghai Jiaotong University. Since October 1992 he has been a research assistant and a Ph.D. candidate at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, Germany. His current research interest is model-based coding.

Instructions to authors

General. Prospective authors are encouraged to submit manuscripts within the scope of the Journal. To qualify for publication, papers must be previously unpublished and not be under consideration for publication elsewhere. All material should be sent in quadruplicate (original plus three copies) to the Editor-in-Chief. Contributors are reminded that once their contribution has been accepted for publication, all further correspondence should not be sent to the Editor, but directly to the publishers (Editorial Department, Elsevier Science B.V., P.O. Box 1991, 1000 BZ Amsterdam, The Netherlands).

All manuscripts will be assessed by at least two (anonymous) referees.

Upon acceptance of an article, the author(s) will be asked to transfer copyright of the article to the publisher. This transfer will ensure the widest possible dissemination of information.

Accepted languages are English (preferred), French and German. The text of the paper should be preceded by abstracts of no more than 200 words in English. Abstracts should contain the substance of the methods and results of the paper. Page proofs will be sent to the principal author with an offprint order form. Fifty offprints of each article can be ordered free of charge. Costs arising from alterations in proof, other than of printer's errors, will be charged to the authors. All pages should be numbered. The first page should include the article title and the author's name and affiliation, as well as a name and mailing address to be used for correspondence and transmission of proofs. The second page should include a list of unusual symbols used in the article and the number of pages, tables and figures. It should also contain the keywords in English.

Figures. All illustrations are to be considered as figures, and each should be numbered in sequence with Arabic numerals. The drawings of the figures must be originals, drawn in black india ink and carefully lettered, or printed on a high-quality laser printer. Each figure should have a caption and these should be listed on a separate sheet. Care should be taken that lettering on the original is large enough to be legible after reduction. Each figure should be identified. The approximate place of a figure in the text should be indicated in the margin. In case the author wishes one or more figures to be printed in colour, the *extra* costs arising from such printing will be charged to the author. In this case 200 offprints may be ordered free of charge. More details are available from the Publisher.

Tables. Tables should be typed on separate sheets. Each table should have a number and a title. The approximate places for their insertion in the text should be indicated in the margin.

Footnotes in text. Footnotes in the text should be identified by superscript numbers and listed consecutively on a separate page.

References. References must be *in alphabetical order* in the style shown below:

Book	[1] A.V. Oppenheim et al., *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1975, Chapter 10, pp. 491–499.
Journal	[2] F.J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform", *Proc. IEEE*, Vol. 66, No. 1, January 1978, pp. 53–83.
Conference Proceedings	[3] D. Coulon and D. Kayser, "A supervised-learning technique to identify short natural language sentence", *Proc. 3rd Internat. Joint Conf. on Pattern Recognition*, Coronado, CA, 8–11 November 1976, pp. 85–89.
Contributed Volume	[4] E.F. Moore, "The firing squad synchronization problem", in: E.F. Moore, ed., *Sequential Machines, Selected Papers*, Addison-Wesley, Reading, MA, 1964, pp. 213–214.

Fast Communications. Papers for the Fast Communications section should be submitted by electronic mail to Prof. Murat Kunt, Laboratoire de Traitement des Signaux, Département d'Electricité, EPFL, Ecublens, CH-1015 Lausanne, Switzerland, Tel.: (4121) 693 26 26, Fax: (4121) 693 46 60, E-mail: fastcom@ltssun17.epfl.ch. Papers should be a maximum of 2,500 words in length (approximately 6 printed journal pages for Signal Processing: Image Communication). Submissions will be subject to the same editorial selection criteria as regular papers. Reviews will be dispatched electronically and decisions will be binary (yes/no) to avoid publication delays. Please ensure your complete postal and e-mail address are indicated on the title page. As no page proofs will be sent to the authors, the presentation should be very clear. For Fast Communications, the figures should be provided in Encapsulated Postscript (eps) format. To ensure fast publication, the manuscript should be written in LaTeX using the document styles of Elsevier Science B.V. Move all files needed (TeX source, eps-files, style and bibliography files) into one directory. Remove all compilation files (*. log, *. lof, *. dvi, *. aux, . . .). Rename the main TeX source into *review.tex* and archive (tar), compress and unencode this directory, and e-mail the unencoded file to: fastcom@ltssun17.epfl.ch. Authors who comply with the above conditions will have their Fast Communication published on the EEE-Alert Server within three weeks of acceptance.

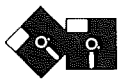
LaTeX files of papers that have been accepted for publication may be sent to the Publisher by e-mail or on a diskette (3.5″ or 5.25″ MS-DOS). If the file is suitable, proofs will be produced without rekeying the text. The article should be encoded in ESP-LaTeX, standard LaTeX, or AMS-LaTeX (in document style 'article'). The ESP-LaTeX package, together with instructions on how to prepare a file, is available from the Publisher. It can also be obtained through the Elsevier WWW home page (http://www.elsevier.nl) or using anonymous FTP from the Comprehensive TeX Archive Network (CTAN). The host-names are: ftp.dante.de, ftp.tex.ac.uk, ftp.shsu.edu; the directory is /tex-archive/macros/latex/contrib/supported/elsevier. *No changes* from the accepted version are permissible, without the explicit approval by the Editors. The Publisher reserves the right to decide whether to use the author's file or not. If the file is sent by e-mail, the name of the journal, *Signal Processing: Image Communication*, should be mentioned in the subject field of the message to identify the paper. Authors should include an ASCII table (available from the Publisher) in their files to enable the detection of transmission errors. The files should be mailed to: Ineke Kolen, Elsevier Science B.V., P.O. Box 103, 1000 AC Amsterdam, The Netherlands. Fax: +31 20 4852829. E-mail: c.kolen@elsevier.nl.

For the purpose of further correspondence the manuscript should end with a complete mailing address, preferably including e-mail address, of at least one of the authors.

ELSEVIER SCIENCE

prefers the submission of electronic manuscripts

Electronic manuscripts have the advantage that there is no need for the rekeying of text, thereby avoiding the possibility of introducing errors and resulting in reliable and fast delivery of proofs.



The preferred storage medium is a 5.25 or 3.5 inch disk in MS-DOS format, although other systems are welcome, e.g. Macintosh.

After **final acceptance,** your disk plus one final, printed and exactly matching version (as a printout) should be submitted together to the accepting editor. **It is important that the file on disk and the printout are identical.** Both will then be forwarded by the editor to Elsevier.

Please follow the general instructions on style/arrangement and, in particular, the reference style of this journal as given in "Instructions to Authors."

Please label the disk with your name, the software & hardware used and the name of the file to be processed.

SIGNAL PROCESSING: IMAGE COMMUNICATION

- ☐ Please send me a free sample copy
- ☐ Please send me subscription information
- ☐ Please send me Instructions to Authors

Name _____

Address _____

_____

_____

Send this coupon or a photocopy to:

ELSEVIER SCIENCE B.V
Attn: Engineering and Technology Department
P.O. Box 1991, 100 BZ Amsterdam, The Netherlands