ESTIMATION OF THE ORIENTATION OF A FACE IN A KNOWLEDGE–BASED ANALYSIS–SYNTHESIS CODER

Liang Zhang , Markus Kampmann

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, D–30167 Hannover, F.R.Germany
Phone: ++49–511–762–5314, email: zhang@tnt.uni–hannover.de

ABSTRACT: *An algorithm for estimation of the orientation of a person's face in a knowledge–based analysis–synthesis coder (KBASC) assuming a talking person appeared in an image sequence is described. For estimation of the orientation of a person's face using a face model, it is assumed that the eye and mouth view lines of a perspective camera are known and the mouth position locates at the right position with respect to its view line. Then, exploiting that eye and mouth positions of the face model along their view lines in the 3D space build an isosceles triangle, eye positions of the face model are iteratively estimated by rotating the eye–mouth plane. For this, the mean square error between the projection of the face model and the real image is minimized using Gauss–Newton–Method. The presented algorithm has been implemented in a knowledge–based analysis–synthesis coder (KBASC) and is applied to improve the tracking accuracy in the KBASC.*

1. INTRODUCTION

For coding of moving images at very low bit rates, an object–based analysis–synthesis coder (OBASC) based on the source model of moving 3D objects has been introduced [1][2]. According to the coding concept, moving real objects are described by moving model objects. Each moving model object is defined by three parameter sets defining its motion, shape and surface color. The parameter sets of each model object are estimated by image analysis algorithms, and no specific knowledge about the scene contents is exploited. If scene specific knowledge is available, an OBASC may be extended to a knowledge–based analysis–synthesis coder (KBASC) [3]. In the case of typical videophone applications, usually a human face appears in the scene, which can be represented by a face model, e.g. the mask *Candide* [4]. Besides automatic adaptation of the face model at the beginning of the image sequence [3][5], tracking a face during the sequence is the main problem [3][6][7][8].

For tracking a face, Kampmann and Ostermann [3], C.S. Choi, K. Aizawa, H. Harashima, T. Takebe [6] and H. Li, R. Forchheimer [7] reported methods based on global motion compensation of a model head. In those algorithms, the 3D motion of the head, i.e. rotation and translation, is estimated and compensated. Because the motion estimation is inaccurate due to noise and inaccurate 3D model head shape, the projections of eye and mouth positions of the face model into the image plane may not match those of the person in the real image. In order to improve the tracking accuracy, Zhang [8] reported a method in which global motion compensation of the head has been extended by an adaptation of the face model to the estimated eyes and mouth positions in the image during the sequence. But this adaptation only takes place in the eye–mouth plane in the 3D space , i.e. the face model can only be shifted and deformed in the eye–mouth plane. The eye–mouth plane in the 3D space is determined by the face model after the 3D motion compensation of the head. Because the accuracy of the 3D motion estimation of the head is affected by the 3D model head shape, the inaccurate 3D model head shape reduces the accuracy of tracking a face.

The orientation of the face model is one of the parameters of the 3D model head shape. Here, the norm of the eye–mouth plane is defined as the orientation of the face model. For the simple face model *Candide*[4], its orientation is only determined by the 3D coordinates of the eyes and the mouth in the 3D space. Thus, the update of the orientation of the face model during the sequence will improve the accuracy of tracking a face.

To estimate the orientation of the person's face, Gee and Cipolla [9] proposed an algorithm which additionally allows rotation of the face model. This algorithm was developed for computer vision applications. The orientation of a face is determined by three ratios of four lengths, e.g. the distance between the *nose tip* and the *nose base*. Those distances were determined by estimation of the locations of facial feature points. Here, limitations in the precision for locating those facial features directly affect the accuracy of the measured distances.

In this contribution, an algorithm is presented which combines the approach in [8] with a strategy for estimation of the orientation of a face using the face model *Candide*. After the approach presented in [8] has been accomplished, it is supposed that the mouth position has already been located at the right position with respect to the view line of a perspective camera. For update of the orientation of the face model, the eye positions are still to be corrected. Presumed that the eye and mouth view lines of a perspective camera are known and exploiting that eye and mouth positions of the face model must not leave their view lines and satisfy an isosceles triangle in the 3D space, eye positions of the face model are iteratively estimated by rotating the eye–mouth plane. For this, the mean square error between the projections of the face model and the real image is minimized using Gauss–Newton–Method. The presented algorithm will be implemented in a knowledge–based analysis–synthesis coder (KBASC) [3]. The improvement of the tracking accuracy by use of this approach is measured by the reduction of the average position errors of the eyes and the mouth of the face model during the sequence.

In this paper, Section 2 addresses the problem of estimation of the orientation of a face using the face model *Candide*. Section 3 discusses the eye positions of the face model along their view lines in the 3D space in the case that the mouth position of the face model is known. Section 4 describes a method for the estimation of the eye positions. This algorithm has been applied to improve the tracking accuracy of a face. Experimental results using real image sequences are given in Section 5.

2. ESTIMATION OF THE ORIENTATION OF A FACE USING THE FACE MODEL *CANDIDE*

For estimation of the orientation of a face using the face model after motion compensation, it is assumed that the difference ΔI in the facial area between the projection of the face model and the real image, except the eye and mouth areas, is due to the error of the orientation of the face model. For the simple face model *Candide*, its orientation is determined by the 3D coordinates of the eyes and the mouth in the 3D space. However, only using a monocular view of a face is not enough to estimate the true positions of the eyes and the mouth of the person in the 3D space. Fortunately, having those true positions is not a necessary condition to update the orientation of the face model. Such a strategy that the mouth position is assumed to be known and the eye positions along their view lines of a perspective camera have to be estimated can be used to achieve this goal. After the approach presented in [8] has been accomplished, it is assumed that the projections of the eyes and the mouth of the face model to the image plane match those positions of the person in the images and the mouth position of the face model locates at the right position with respect to its view line. Furthermore, the 3D coordinates of the eyes and the mouth of the face model build an isosceles triangle in the 3D space, i.e.,

$$(\boldsymbol{P}^l - \boldsymbol{P}^r) \cdot (\boldsymbol{P}^m - \boldsymbol{P}^o) = 0\,, \qquad (1)$$

where $\boldsymbol{P} = (P_x, P_y, P_z)$ is a 3D point in the model world and superscripts l, r, m and o stand for the left eye, the right eye, the mouth and the center of both

eyes, respectively. Therefore, estimation of the orientation can be formulated as follows,

$$\sum_{\substack{\text{all pels in the facial area} \\ \text{except the eye and mouth areas}}} (\Delta I)^2 =$$

$$\sum_{\substack{\text{all pels in the facial area} \\ \text{except the eye and mouth areas}}} (I_s - I_m(\boldsymbol{P}^l, \boldsymbol{P}^r))^2 \rightarrow MIN, \quad (2)$$

with conditions Eq.(1),

$$\boldsymbol{P}^m = const., \quad (3)$$

and

$$\boldsymbol{P}^l, \boldsymbol{P}^r \in \Re, \quad (4)$$

i.e., the orientation of a face using the face model is estimated in such a way that the mean square error in the facial area between the projection of the face model $I_m(\boldsymbol{P}^l, \boldsymbol{P}^r)$ and the real image I_s is minimized. Since the eye 3D coordinates $\boldsymbol{P}^l = (P^l_x, P^l_y, P^l_z)$ and $\boldsymbol{P}^r = (P^r_x, P^r_y, P^r_z)$ are estimated along their view lines of a perspective model camera, $\boldsymbol{P}^l$ and $\boldsymbol{P}^r$ can be determined if the z–coordinate P^l_z and P^r_z are known. Therefore, estimation of the orientation is equal to estimation of the z–coordinates of both eyes of the face model, i.e.,

$$\sum_{\substack{\text{all pels in the facial area} \\ \text{except the eye and mouth areas}}} (I_s - I_m(P^l_z, P^r_z))^2 \rightarrow MIN. \quad (5)$$

Due to the geometrical condition (1), the position of the left eye is dependent on that of the right eye and *vice versa*. This problem is addressed in the next Section.

3. EYE POSITIONS OF THE FACE MODEL ALONG THEIR VIEW LINES OF A PERSPECTIVE CAMERA IN THE 3D SPACE

In order to simplify the following discussions, symbols θ_l and θ_r are introduced, 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. In the case that the 3D mouth coordinate of the face model is known and the values of the 3D eye coordinates must be real, three cases for real 3D–coordinates of both eyes of the face model along their view lines of a perspective camera are deduced from Eq.(1):

- If $\theta_l > \theta_r$, there is no constraint for the left eye. But for the right eye, not any value along the view line can be selected as its position. It means that there is an interval along the view line for the right eye, in which it is not allowed to be located;
- If $\theta_l < \theta_r$, there is no constraint for the right eye. But for the left eye, not any value along the view line can be selected as its position. It means that there is an interval along the view line for the left eye, in which it is not allowed to be located;
- If $\theta_l = \theta_r$, there is no constraint for both eyes. It means that any real value can be taken as the positions of both eyes along their view lines.

4. ALGORITHM FOR ESTIMATION OF THE Z–COORDINATES OF BOTH EYE POSITIONS OF THE FACE MODEL

In this Section, an algorithm for estimation of the z–coordinates of both eye positions of the face model is presented. As discussed in Section 2, estimation of the orientation is equal to estimation of the z–coordinates (P^l_z, P^r_z) of both eyes of the face model, whereas (P^l_z, P^r_z) is estimated in such a way that the mean square error between the projection of the face model $I_m(P^l_z, P^r_z)$ and the real image I_s is minimized. For minimization, a Gauss–Newton–Method is applied which uses a set of observation points from the face model. Each observation point $\boldsymbol{O}^{(j)}_k = (\boldsymbol{P}^{(j)}_k, \boldsymbol{g}^{(j)}, I^{(j)})$ at time instant k is located on the face model surface and is described by its 3D position $\boldsymbol{P}^{(j)}_k = (P^{(j)}_x, P^{(j)}_y, P^{(j)}_z)^T$, its luminance value $I^{(j)}$ and its linear gradients $\boldsymbol{g}^{(j)} = (g^{(j)}_x, g^{(j)}_y)^T$. The luminance and linear gradients are taken from the same image from which the color parameters of the 3D model object were derived. The measure for selecting observation points is a high spatial gradient. For each observation point $\boldsymbol{O}^{(j)}_k$, the luminance difference $\Delta I^{(j)}$ between the projections of the face model and the real image is related to z–coordinates of the eyes by the following linearized equation:

$$\Delta I^{(j)} = \frac{\partial \Delta I^{(j)}}{\partial P^l_z} \cdot \Delta P^l_z + \frac{\partial \Delta I^{(j)}}{\partial P^r_z} \cdot \Delta P^r_z \quad (6)$$

where $\frac{\partial \Delta I^{(i)}}{\partial P^l_z}$ and $\frac{\partial \Delta I^{(i)}}{\partial P^r_z}$ are only depending on the observation point $\boldsymbol{O}^{(j)}_k$ and the eye and mouth positions $(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m)$ before update of the orientation of a

face, and $(\Delta P_z^l, \Delta P_z^r)$ are the alterations of the z–coordinates of the eyes.

In order to satisfy Eq. (1), the eye positions $(\boldsymbol{P}^l, \boldsymbol{P}^r)$ are dependent on each other. From Eq. (1), there is

$$\Delta P_z^r = R(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m) \cdot \Delta P_z^l, \tag{7}$$

where $R(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m)$ is a coefficient only depending on the eye and mouth positions $(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m)$ before update of the orientation of a face. According to the results of Section 3, substituting Eq.(7) into Eq. (6), we have

- $\theta_l > \theta_r$ **:**

$$\Delta I^{(j)} = \left[\frac{\partial \Delta I^{(j)}}{\partial P_z^l} + \frac{\partial \Delta I^{(j)}}{\partial P_z^r} \cdot R(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m) \right] \cdot \Delta P_z^l \ , \tag{8}$$

- $\theta_l < \theta_r$ **or** $\theta_l = \theta_r$ **:**

$$\Delta I^{(j)} = \left[\frac{1}{R(\boldsymbol{P}^l, \boldsymbol{P}^r, \boldsymbol{P}^m)} \cdot \frac{\partial \Delta I^{(j)}}{\partial P_z^l} + \frac{\partial \Delta I^{(j)}}{\partial P_z^r} \right] \cdot \Delta P_z^r. \tag{9}$$

The residuum of the equation system in the different cases is then minimized by a Gauß method for least squares error,

$$\sum_{\boldsymbol{o}^{(j)}} (\Delta I^{(j)})^2 \rightarrow MIN. \tag{10}$$

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

- $\theta_l > \theta_r$**:**

$$P_z^{'l} = P_z^l - \Delta P_z^l \ , \tag{11}$$

- $\theta_l < \theta_r$ **or** $\theta_l = \theta_r$ **:**

$$P_z^{'r} = P_z^r - \Delta P_z^r \ , \tag{12}$$

whereas the z–coordinate of the other eye is calculated based on (1). However, Eq.(1) usually gives two values of the eye z–coordinate. Thus, one value with the smaller mean square error between the projections of the face model and the real image will be selected as the eye z–coordinate. Then, the x– and y–coordinates of both eyes are calculated according to the principle of the projection of a perspective model camera. Afterwards, a new set of estimation equations is established, which give new z–coordinates of the eyes updates. Due to the linearization, the z–coordinates of the eyes have to be estimated iteratively. In case of convergency, the z–coordinate updates approach zero during the iterations.

5. EXPERIMENTAL RESULTS

The described algorithm has been applied to track a face in natural sequences, i.e. the *Claire* and *Miss America* sequences, with a spatial resolution corresponding to CIF and a frame rate of 10 Hz. In order to measure the tracking accuracy of a face in the KBASC, the average position errors

$$\bar{f}_{eye} = \frac{1}{N_{eye}} \sum_{n=1}^{N_{eye}} \max[|X_l - X_l^t|, |Y_l - Y_l^t|, |X_r - X_r^t|, |Y_r - Y_r^t|] \tag{13}$$

$$\bar{f}_{mouth} = \frac{1}{N} \sum_{n=1}^{N} \max[|X_m - X_m^t|, |Y_m - Y_m^t|] \tag{14}$$

are introduced, where (X_l^t, Y_l^t), (X_r^t, Y_r^t) and (X_m^t, Y_m^t) are the true image coordinates of the left open eyes, the right open eyes and the mouth in the image sequence (manually determined), (X_l, Y_l), (X_r, Y_r) and (X_m, Y_m) are the coordinates of the left open eyes, the right open eyes and the mouth of the face model projected onto the image plane. N is the number of images. N_{eye} is the number of images in which the eyes are open. Figs. 1 and 2 show the results for both test sequences. Compared to the algorithm presented in [8], the method proposed here reduces the average position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ by 0.32 and 0.02 pel for the test sequence *Claire,* and by 0.1 and 0.36 pel for the test sequence *Miss America*, respectively.

6. CONCLUSION

In this contribution an algorithm for estimation of the orientation of a person's face in a knowledge–based analysis–synthesis coder (KBASC) is proposed. By application of this method in face tracking, the average position errors between the true image coordinates of eyes and mouth and the projected coordinates of eyes and mouth of the face model onto the image plane are reduced for both test sequences *Claire* and *Miss America* by 15 percent for the eyes and by 10 percent for the mouth, respectively.

7. REFERENCES

[1] J. Ostermann, "Object–based analysis–synthesis based on the source model of moving

rigid 3D objects", Signal Processing: Image Communication, Vol. 6, No. 2, May 1994, pp. 143–161.

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

[3] M. Kampmann, J. Ostermann "Automatic adaptation of a face model in a layered coder with an object–based analysis–synthesis layer and a knowledge–based layer", Accepted for publication in Signal Processing: *Image Communication.*

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

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

[6] C.S. Choi, K. Aizawa, H. Harashima, T.Takebe, "Analysis and synthesis of facial image sequences in model–based image coding", IEEE Trans. on Circuits and Systems for Video Technology, Vol.4, No. 3, June 1994, pp.257–275.

[7] H. Li, R. Forchheimer, "Two–view facial movement estimation", IEEE Trans. on Circuits and Systems for Video Technology, Vol.4, No. 3, June 1994, pp.276–287.

[8] L. Zhang, "Tracking a face in a knowledge–based analysis–synthesis coder", International Workshop on Coding Techniques for Very Low Bit–rate Video (VLBV 95), Tokyo, Japan, November 1995, No. A.6.

[9] A. Gee and R. Cipolla, "Determining the gaze of faces in images", Image and Vision Computing, Vol.12, No.10, December 1994, pp.639–647.

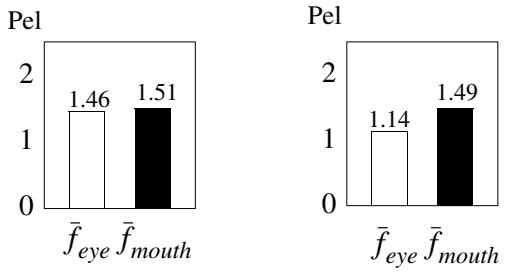

Fig. 1 Average position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ for the test sequence "Claire" (CIF, 10Hz)
(a) tracking only with the algorithm in [8]
(b) tracking with the proposed algorithm

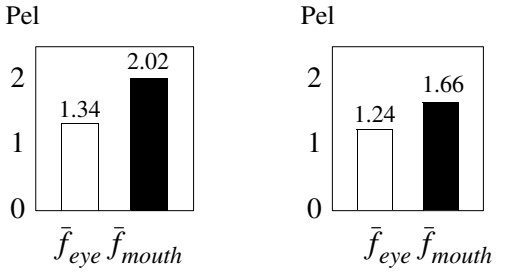

Fig. 2 Average position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ for the test sequence "Miss America" (CIF,10Hz)
(a) tracking only with the algorithm in [8]
(b) tracking with the proposed algorithm