TRACKING A FACE IN A KNOWLEDGE–BASED ANALYSIS–SYNTHESIS CODER

Liang Zhang

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 tracking a face in a knowledge–based analysis–synthesis coder (KBASC) is described. After motion compensation of the head, eye and mouth center positions of a person in an image are estimated. Then, the position of the face model is corrected using the estimated positions of eyes and mouth. Experimental results show that the proposed algorithm allows a more accurate tracking of a face in a KBASC compared to tracking only with motion compensation of the head.

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 model objects has been introduced[1][2]. According to the coding concept, real objects are described by model objects. Each moving model object of an image is described by three parameter sets defining its motion, shape and surface color. The parameter sets of each model object are estimated by an image analysis, and no 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, which can be represented by a face model, e.g. the mask *Candide*[4], appears in the scene. Besides automatic adaptation of the face model[3], tracking is the main problem.

After adaptation, the face model is tracked together with the person's moving head by global motion compensation [3][5][6]. The motion of the head, i.e. rotation and translation, is estimated and compensated. Because the motion estimation is inaccurate due to noise, facial expression and inaccurate 3D–model shape, the projections of eye and mouth positions of the face model may not match those of the person in the real image.

In this contribution, an extended algorithm for tracking a face in a knowledge–based analysis–synthesis coder is presented. As a first stage of the extended algorithm, the motion compensation presented in [3] is used. After motion compensation, the center positions of eyes and mouth of the person in the image sequence are estimated. Then, the location of the face model is corrected using the estimated positions of eyes and mouth, if the deviation between the estimated positions and the projections of eye and mouth positions of the face model onto the image plane exceeds a given threshold.

Section 2 describes the estimation of eye and mouth center positions in the image sequence. Section 3 explains the correction of the position of the face model. Experimental results using real image sequences are given in Section 4.

2. Estimation of eye and mouth center positions in an image

An approach for the estimation of eye and mouth center positions is presented in [3] only for the adaptation of a face model once at the start of the sequence. Here, this approach is also used for tracking during the whole sequence and extended in several respects in order to achieve a reliable and more precise estimation of eye and mouth center positions.

In order to estimate the correct center positions of the eyes and the mouth in the real image, large search areas are used in [3]. In the case of tracking, after motion compensation of the head 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, in the presented approach the projected positions of eyes and mouth of the face model onto the image plane are exploited. By this, the correct center positions of eyes and mouth can be estimated more often and the search areas can be essentially diminished.

In the search areas for eyes and mouth, the correlation between the real image and templates for eyes and mouth created in the same way as the templates in [3] is computed. The templates are roughly adapted to the size of the person's face. Compared to [3], in this approach the templates are additionally inclined in accordance with the inclination of the head.

Areas with high correlation between the real image and the eye templates are extracted and the two darkest points in these areas are selected as eye center positions, assuming that the eyes are open. The point with the highest correlation in the mouth search area is selected as mouth center position.

3. Correction of the position of the face model in the eye–mouth plane

After the center positions of eyes and mouth have been estimated, the deviation d_p between the estimated positions and the projections of eye and mouth center positions of the face model onto the image plane is measured as:

$$d_p = \max(|X_l - X_l'|, |Y_l - Y_l'|, |X_r - X_r'|, |Y_r - Y_r'|, |X_m - X_m'|, |Y_m - Y_m'|), \quad (1)$$

where (X_l, Y_l), (X_r, Y_r) and (X_m, Y_m) are the estimated positions of the left eye, the right eye and the mouth, respectively, and (X_l', Y_l'), (X_r', Y_r') and (X_m', Y_m') are the image coordinates of eyes and mouth of the face model. Depending on d_p and two given thresholds T_{h_1} and T_{h_2}, the face model's position is corrected only if $T_{h_1} < d_p < T_{h_2}$. For $d_p \leq T_{h_1}$, the position of the face model is accurate enough. For $d_p \geq T_{h_2}$, the position of the face model is not corrected, because the positions of eyes and mouth are not reliable. Also, if the center positions of the eyes could not at all be estimated, e.g. if the eyes are closed, no correction is performed and only motion compensation of the head with the face model is applied.

The position of the face model *Candide*[4] is determined only by the 3D–coordinates of eyes and mouth. Therefore, the new 3D–coordinates of eyes and mouth of the face model must be calculated in order to correct the position of the face model. Since the correct depths of eyes and mouth are unknown, the calculation of the new 3D–coordinates of eyes and mouth is based on the following two conditions:

1 3D–coordinates of eyes and mouth are in the same plane (eye–mouth plane) before and after the correction of a face model's position.
2 3D–coordinates of eyes and mouth span an isosceles triangle.

The projections of the center positions of eyes and mouth estimated in Section 2 from the image plane onto the eye–mouth plane of the face model may not fulfill the second condition. Because of a more precise estimation of the eye center positions than of the mouth center position in the image plane, the projections of the estimated eye positions from the image plane onto the eye–mouth plane are taken as the new 3D eye coordinates. According to condition 2, the new 3D mouth position lies on the perpendicular bisection of the line between both new 3D eye positions. This line is projected onto the image plane. Here, it intersects a parallel of the line connecting both estimated eye positions in the image plane. This parallel goes through the estimated mouth position in the image plane. The projection of this intersection point from the image plane onto the eye–mouth plane gives the new 3D mouth coordinates.

After the new 3D coordinates of eyes and mouth are determined, the face model is scaled again und shifted to the new position.

4. Experimental results

The described algorithm was applied to the test sequences *Claire* and *Miss America* 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 a KBASC, the position errors

$$\bar{f}_{eye} = \frac{1}{N_{eye}} \sum_{n=1}^{N_{eye}} \max\left[|X_l' - X_l^t|, |Y_l' - Y_l^t|, |X_r' - X_r^t|, |Y_r' - Y_r^t|\right], \tag{2}$$

$$\bar{f}_{mouth} = \frac{1}{N} \sum_{n=1}^{N} \max\left[|X_m' - X_m^t|, |Y_m' - Y_m^t|\right], \tag{3}$$

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 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 image coordinates of eyes and mouth of the face model. N_{eye} is the number of images in which the eyes are open. N is the number of images. Fig. 1 and Fig.2 show the results for the test sequences *Claire* and *Miss America*, respectively. The thresholds $T_{h_1} = 2$ pels and $T_{h_2} = 5$ pels have been chosen for both test sequences.

For the test sequence *Claire,* the proposed algorithm reduces the position errors $\bar{f}_{mouth}$ by 1.11 pel compared to tracking only with motion compensation of the head [3]. The position error $\bar{f}_{eye}$ is nearly the same for the proposed algorithm and for tracking only with motion compensation. For the test sequence *Miss America,* the proposed algorithm reduces the position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ by 0.36 and 1.3 pel respectively compared to tracking only with motion compensation.

5. Conclusion

In this contribution an algorithm for tracking a face in a knowledge–based analysis–synthesis coder (KBASC) is proposed. After motion compensation of the face model, the center positions of eyes and mouth in the image are estimated. According to the estimated positions

of eyes and mouth, the location of the face model is corrected. For the test sequences *Claire* and *Miss America*, the average deviation 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 is reduced for both test sequences. In the future the orientation of the eye–mouth plane of the face model will be taken into consideration in order to further improve tracking of a face.

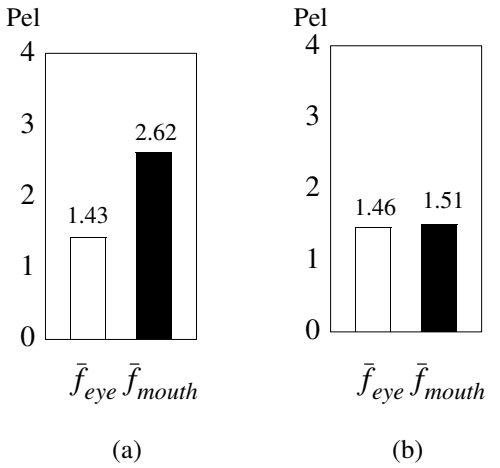

Fig. 1 Position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ for test sequence "Claire" (CIF, 10Hz)
(a) tracking only with motion compensation of the head [3]
(b) tracking with the proposed algorithm

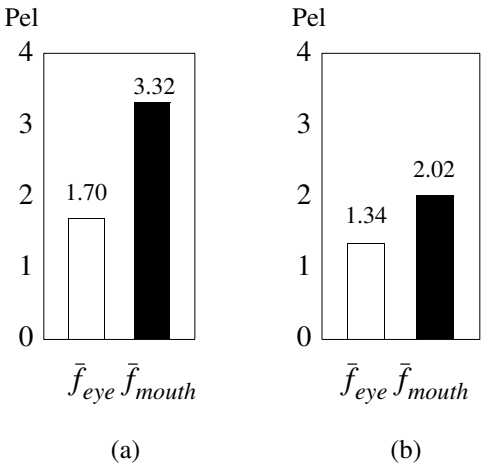

Fig. 2 Position errors $\bar{f}_{eye}$ and $\bar{f}_{mouth}$ for test sequence "Miss America" (CIF, 10Hz)
(a) tracking only with motion compensation of the head [3]
(b) tracking with the proposed algorithm

References

[1] J. Ostermann, "Object–based analysis–synthesis based on the source model of moving rigid 3D objects", Signal Processing: Image Communication, Vol. 6, pp. 143–161, 1994

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

[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", submitted for publication to Signal Processing: Image Communication, Special Issue on Communicating Natural Scenes by Image Synthesis.

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

[5] 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, pp.257–275, 1994

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