Automatic Adaptation of a Face Model Using Action Units

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, http://www.tnt.uni–hannover.de/∽zhang

ABSTRACT: *Automatic adaptation of a face model is one of the main topics for semantic coding of videophone sequences. To implement mimic analysis by means of action units in subsequent frames, the initial values for the intensity parameters of action units have to be determined. Up to now, no action unit is exploited for the adaptation. In this paper, an adaptation algorithm using action units, which consists of three steps, namely: (i) global adaptation, (ii) local adaptation, (iii) mimic adaptation, is proposed. The experimental results obtained with typical videophone sequences show that the face model Candide matches the eye and mouth features very well.*

1. INTRODUCTION

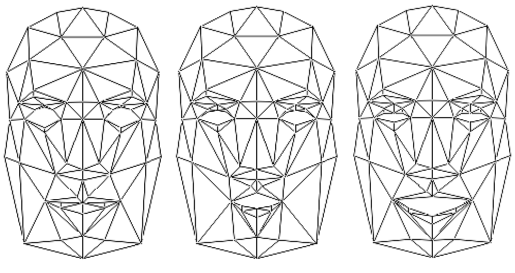
For coding of head–and–shoulder videophone sequences at very low bit rates, semantic coding (also called model–based coding) has been proposed[1][12]. In this coding scheme, the face of the person in the scene is represented by a generic 3D face model, e.g. *Candide* [9], and its facial expressions are described by mimic parameters, e.g. muscle parameters[4] or action units[3]. The mimic parameters are estimated at the transmitter, coded and transmitted. At the receiver, a model image can be synthesized using facial animation. At the beginning of the image sequence, the face model has to be adapted to the individual face in the scene. For this, the size and the position of the face model have to be adapted. Furthermore, the mimic of the face model also has to be adapted and the initial values for the intensity parameters of action units have to be determined.

For adaptation of a face model, some approaches have been proposed. In [1][12], a face model is adapted manually. In [6], the face model *Candide* is adapted only using the center points of the eyes and the mouth. In [11], a 3D generic face model onto which a generic facial texture has already been mapped by hand before is adapted to a person's face in the scene by a steepest–gradient search method. In [5], an algorithm adapts the face model from [11] to the facial features. No initial values for the muscle parameters of the face model have been determined. In [7], an algorithm adapts the face model *Candide* to the facial features in two steps. The first step is global adaptation which consists of 3D rotation, translation and scaling of the face model. The second step is local adaptation which provides a more refined fit to the contours of the real face by estimating a displacement vector for each vertex of the face model. In [8], an algorithm adapts the face model to the facial features by first moving the control vertices, which correspond to facial features, to the feature points of the face and secondly shifting all the other vertices according to the movement of the control vertices in the first step. However, none of these algorithms explots mimic parameters for adaptation and no initial values of the mimic parameters for mimic analysis are determined.

In this contribution, an algorithm for the automatic adaptation of the face model *Candide* to the eye and mouth features using action units is proposed. The adaptation algorithm consists of three steps. The first step is global adaptation which positions and scales the face model in the 3D model world. The second step is local adaptation in which the eye corner points and the lip's thickness of the face model are adapted. The final step is mimic adaptation in which the eye and mouth openings as well as the mouth corner positions are adapted using six action units.

The six action units [3] used for mimic adaptation are listed in Table 1: two action units for the eyes and four action units for the mouth. For simplification, an action unit with an intensity parameter k is abbreviated as AUi(k) where i is the number of the action unit. Two examples for combining different action units are shown in Fig. 1b and 1c.

This contribution is organized as follows: following the introduction, the estimation of facial features is briefly addressed in Section 2. Section 3 describes the adaptation algorithm. Some experimental results obtained with typical natural head–and–shoulder videophone sequences are given in Section 4. Conclusions can be found in Section 5.



(a) (b) (c)

Fig. 1 Face model *Candide* [9]:
(a) Generic face model;
(b) Face model with AU41(0.2), AU7(0.3), AU8(0.5), AU10(0.6) and AU16(0.2);
(c) Face model with AU41(0.4), AU7(0.6), AU12(0.5), AU10(0.3) and AU16(0.5).

2. ESTIMATION OF FACIAL FEATURES

For adaptation of a face model, facial features of the human face in the scene have to be automatically estimated. Thus, this section briefly reviews how to estimate facial features.

For estimation of facial features, many algorithms have been proposed and overviews of this topic can be found in [2][10]. Among all facial features, the eye and mouth features are the most important. For estimation of these features, an approach using deformable templates as proposed in [13] has been developed in [14][15][16] [17]. There, the eye and mouth features are described by deformable templates and represented by the center and corner points of the eyes and the mouth as well as the eyelid and the lip outline parameters. The eyelid outline parameters describe the opening height of the eyes. The lip outline parameters denote the opening height of the mouth as well as the thickness of the lips. With accurate face tracking [15], the projections of the face model *Candide* is exploited to give search windows for the eye and mouth features. The center and corner points of the eyes and the mouth are first estimated by textured template matching [14][15]. Afterwards, minimizing a cost function gives the estimates of the eyelid and the lip outline parameters [16][17]. This algorithm has been applied to the sequences *Akiyo*, *Claire* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz. Some results are shown in Fig. 4b.

3. ADAPTATION OF THE FACE MODEL *CANDIDE*

The whole adaptation algorithm consists of three steps, namely: (i) the global adaptation; (ii) the local adaptation; (iii) the mimic adaptation. These three steps are addressed in the following subsections.

Global adaptation

The global adaptation of a face model consists of scaling, rotation and translation of the face model in the 3D model world. Let an arbitrary point $\boldsymbol{P}^{*(i)}$ on the surface of a face model move to its new position $\boldsymbol{P}^{'(i)}$ in the 3D model world. Then, this global adaptation can be formulated as

$$\boldsymbol{P}^{'(i)} = [\boldsymbol{R}_0] \cdot [\boldsymbol{S}] \cdot (\boldsymbol{P}^{*(i)} - \boldsymbol{C}^0) + \boldsymbol{T} \qquad (1)$$

with the translation vector $\boldsymbol{T}$, the face model center $\boldsymbol{C}^0$, the rotation matrix $[\boldsymbol{R}_0]$, and the scale factor matrix

$$[\boldsymbol{S}] = \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & S_z \end{bmatrix}, \qquad (2)$$

Table 1. Action units used in this contribution

Action unit name	**Intensity parameter k**	**Number of the included vertices**	**Abbr.**
Lid Droop	0 to 1	6	AU41(k)
Lid Tightener	0 to 1	2	AU7(k)
Lips Toward	0 to 1	12	AU8(k)
Lip Corner Puller	0 to 1	6	AU12(k)
Upper Lip Raiser	0 to 1	7	AU10(k)
Lower Lip Depressor	0 to 1	9	AU16(k)

where S_x, S_y, and S_z are the scale factors along the x–, y– and z–axis. The face model center $\boldsymbol{C}^0$ is here defined as the center between both eyes of the face model. For the face model *Candide,* these parameters are determined only by the 3D center eye and mouth positions. Therefore, the global adaptation is only a problem of how to estimate the 3D eye and mouth center positions in the 3D model world from the estimated 2D feature points in the image.

The 2D eye and mouth centers are here defined as the centers of the eye and mouth corner positions which are estimated as described in [14]. Since the true depths of the eyes and the mouth are unknown, some assumptions have to be made for the calculation of these 3D center positions of the face model in the 3D model world. With face tracking [15], the face model has already been matched to the real face and this position of the face model can be exploited to calculate the 3D center positions of the face model in the 3D model world. Therefore, it is assumed that the required 3D eye and mouth center positions are in the eye–mouth–plane. The eye–mouth–plane is here defined by the 3D center positions of the eyes and mouth of the face model after face tracking [15].

Let the 3D positions $\boldsymbol{P}^l$, $\boldsymbol{P}^r$, $\boldsymbol{P}^m$ and $\boldsymbol{P}^o$ be projections of the 2D center positions of the left eye, the right eye, the mouth and the center between both eyes from the image plane onto the eye–mouth–plane (Fig. 2). Then, the angle α between the line $\boldsymbol{P}^o\boldsymbol{P}^l$ and the line $\boldsymbol{P}^o\boldsymbol{P}^m$ is calculated. If α is equal to 90^o, i.e. the projections of the 2D center positions of the eyes and the mouth from the image plane onto the eye–mouth–plane

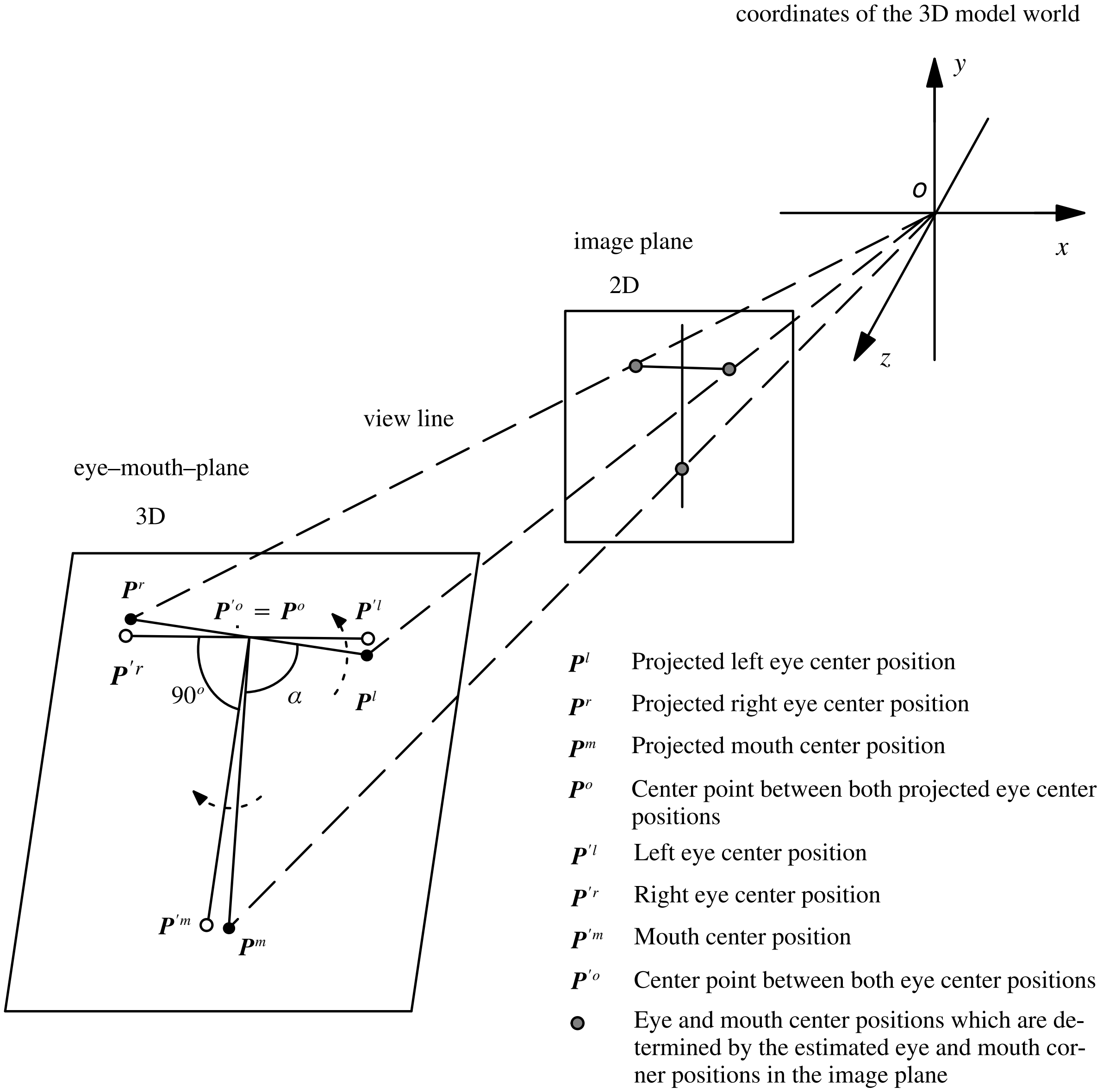


Fig. 2 Calculation of the 3D eye and mouth center positions of the face model. α is the angle between line $\boldsymbol{P}^o\boldsymbol{P}^l$ and $\boldsymbol{P}^o\boldsymbol{P}^m$.

span an isosceles triangle, these projected 3D eye and mouth positions are the required 3D eye and mouth center positions $\boldsymbol{P}'^{l}$, $\boldsymbol{P}'^{r}$, $\boldsymbol{P}'^{m}$ and $\boldsymbol{P}'^{o}$ of the face model.

Since the projections of the 2D center positions of the eyes and the mouth from the image plane onto the eye–mouth–plane may not always span an isosceles triangle, e.g. $\alpha < 90^o$, the projected 3D eye and mouth center positions have to be shifted such that they span an isosceles triangle. In this case, the criterion for calculating the 3D eye and mouth center positions is that the calculated 3D eye and mouth center positions have the same accuracy, i.e. they must fulfill the conditions

$$\overline{\boldsymbol{P}^o\boldsymbol{P}^l} \cdot \measuredangle \boldsymbol{P}'^l\boldsymbol{P}^o\boldsymbol{P}^l = \overline{\boldsymbol{P}^o\boldsymbol{P}^m} \cdot \measuredangle \boldsymbol{P}'^m\boldsymbol{P}^o\boldsymbol{P}^m \tag{3}$$

and

$$|\measuredangle \boldsymbol{P}'^l\boldsymbol{P}^o\boldsymbol{P}^l| + |\measuredangle \boldsymbol{P}'^m\boldsymbol{P}^o\boldsymbol{P}^m| = |90^o - \alpha| \,. \tag{4}$$

From Eq. (3) and (4), we have

$$|\measuredangle \boldsymbol{P}'^l\boldsymbol{P}^o\boldsymbol{P}^l| = |90^o - \alpha| \cdot \frac{\overline{\boldsymbol{P}^o\boldsymbol{P}^m}}{\overline{\boldsymbol{P}^o\boldsymbol{P}^l} + \overline{\boldsymbol{P}^o\boldsymbol{P}^m}} \tag{5}$$

and

$$|\measuredangle \boldsymbol{P}'^m\boldsymbol{P}^o\boldsymbol{P}^m| = |90^o - \alpha| \cdot \frac{\overline{\boldsymbol{P}^o\boldsymbol{P}^l}}{\overline{\boldsymbol{P}^o\boldsymbol{P}^l} + \overline{\boldsymbol{P}^o\boldsymbol{P}^m}} \,. \tag{6}$$

If $\alpha < 90^o$, rotating counterclockwise the line $\overline{\boldsymbol{P}^r\boldsymbol{P}^l}$ by $\measuredangle \boldsymbol{P}'^l\boldsymbol{P}^o\boldsymbol{P}^l$ and clockwise the line $\overline{\boldsymbol{P}^o\boldsymbol{P}^m}$ by $\measuredangle \boldsymbol{P}'^m\boldsymbol{P}^o\boldsymbol{P}^m$ around the position $\boldsymbol{P}^o$ in the eye–mouth–plane give the required 3D eye and mouth center positions $\boldsymbol{P}'^{l}$, $\boldsymbol{P}'^{r}$, $\boldsymbol{P}'^{m}$ and $\boldsymbol{P}'^{o}$ of the face model. Otherwise, rotating clockwise the line $\overline{\boldsymbol{P}^r\boldsymbol{P}^l}$ by $\measuredangle \boldsymbol{P}'^l\boldsymbol{P}^o\boldsymbol{P}^l$ and counterclockwise the line $\overline{\boldsymbol{P}^o\boldsymbol{P}^m}$ by $\measuredangle \boldsymbol{P}'^m\boldsymbol{P}^o\boldsymbol{P}^m$ around the position $\boldsymbol{P}^o$ in the eye–mouth–plane give the required 3D eye and mouth center positions $\boldsymbol{P}'^{l}$, $\boldsymbol{P}'^{r}$, $\boldsymbol{P}'^{m}$ and $\boldsymbol{P}'^{o}$ of the face model.

After the required 3D eye and mouth center positions of the face model in the 3D model world have been determined, all parameters in Eq. (1) can be calculated. Afterwards, the face model is scaled and shifted in the 3D model world according to Eq. (1).

Local adaptation

With the global adaptation, the centers of the eyes and the mouth of the face model have been matched to their real positions in the image. The local adaptation of the face model *Candide,* which takes place after the global adaptation as described above, aims at a more refined fit to the eye corner positions and the lip thickness.

For the adaptation of the eye corner positions, the eyes of the face model, which include 10 vertices each (Fig. 3), are locally scaled and inclined so that the projections of the eye corner positions of the face model onto the image plane are matched to the estimated 2D eye corner positions. Let $\overline{L_{eye}}$ be the average width of both real eyes which is calculated according to the estimated eye corner positions. Let $L_{eye\ model}$ be the eye width of the face model which is projected onto the image plane. Then, the local scale factor S is calculated as

$$S = \frac{\overline{L_{eye}}}{L_{eye\ model}} \,. \tag{7}$$

For the inclination of each eye, let $\boldsymbol{p}_r$ and $\boldsymbol{p}_l$ be the estimated right and left corner positions of an eye in the image plane, respectively. Then, the inclination of the eye is defined as

$$\beta = \arctan\left(\frac{\boldsymbol{p}_l(Y) - \boldsymbol{p}_r(Y)}{\boldsymbol{p}_l(X) - \boldsymbol{p}_r(X)}\right) , \tag{8}$$

where $(X,\ Y)$ are the coordinates in the image plane. With the local scale factor S and the inclination β, all vertices of each eye of the face model are first scaled and then inclined.

For the adaptation of the thickness of the lips, the new 3D coordinates of its control vertex of the face model are first calculated according to the estimated lip thckness. After that, they are shifted to these new positions.

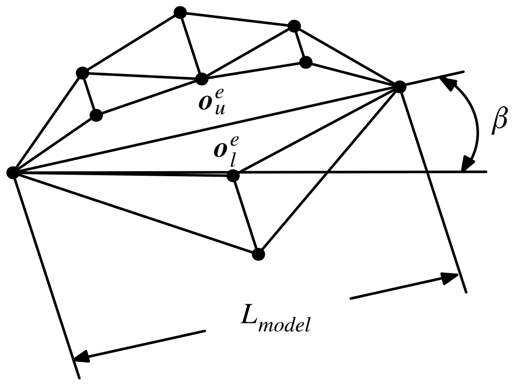


Fig. 3 Vertices of an eye model. L_{model} is the width of the eye. β is the inclination of the eye in the image plane. The points $\boldsymbol{o}_u^e$ and $\boldsymbol{o}_l^e$ are the control vertices of the opening of the eye;

Mimic adaptation

After the global and the local adaptation of the face model have been carried out, the eye and mouth openings as well as the mouth corner positions of the face model have to be adapted. For this adaptation, six action units, which are listed in Table 1, are exploited. Let an arbitrary point $\boldsymbol{P}''^{(i)}$ on the surface of the face model after the global and the local adaptation move to its new position $\boldsymbol{P}'''^{(i)}$ in the 3D model world after the mimic adaptation. Thus, the mimic adaptation can be formulated as

$$P'''^{(i)} = P''^{(i)} + \sum_i AUi(k_i) . \tag{9}$$

Therefore, the main problem of the mimic adaptation is how to estimate the intensity parameter k_i for each action unit AUi.

For the opening height of the eyes, two action units AU41 and AU7 are used to adapt the opening height of the eyes of the face model. Let $\boldsymbol{o}_u^e$ and $\boldsymbol{o}_l^e$ be the projections of the control vertices of the opening height of the upper and lower eyelid (Fig. 3), respectively. Here, the superscript e stands for the eye, subscripts u and l stand for upper and lower eyelid respectively. For the upper eyelid, the opening height is defined as the distance $D_{o_u^e}$ from the point $\boldsymbol{o}_u^e$ vertically to the line connecting both projections of the eye corner positions of the face model. For the lower eyelid, the opening height is defined as the distance $D_{o_l^e}$ from the control point $\boldsymbol{o}_l^e$ vertically to the line connecting both projections of the eye corner positions of the face model. For $k_i = 0$ (i = 41, 7), the opening heights of the eyelids of the face model are

$$D_{o_u^e} = D_{o_u^e}^{\max} \tag{10}$$

and

$$D_{o_l^e} = D_{o_l^e}^{\max} , \tag{11}$$

where $D_{o_u^e}^{\max}$ and $D_{o_l^e}^{\max}$ are the maximum values of $D_{o_u^e}$ and $D_{o_l^e}$. For $k_i = 1$ (i = 41, 7), the eyes are closed.

Let d_u^e and d_l^e be the estimated opening heights of the eyelids in the image plane. Thus, the intensity parameters k_i (i = 41, 7) are calculated as

$$k_{41} = \frac{D_{o_u^e}^{\max} - d_u^e}{D_{o_u^e}^{\max}} \tag{12}$$

and

$$k_7 = \frac{D_{o_l^e}^{\max} - d_l^e}{D_{o_l^e}^{\max}} . \tag{13}$$

With the estimated intensity parameters k_{41} and k_7, the face model is first adapted by

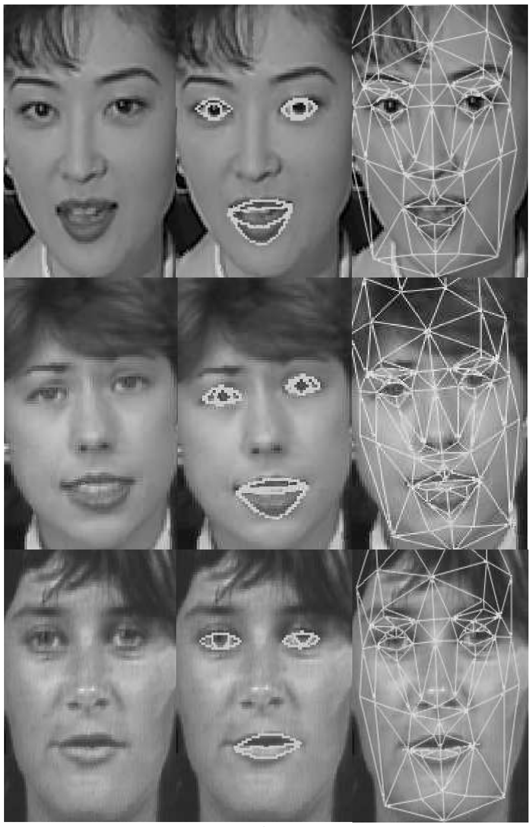
$$P'''^{(i)} = P''^{(i)} + \sum_i AUi(k_i) . \tag{14}$$

After that, the intensity parameters k_i (i = 8, 12, 10, 16) can be also estimated in the similar way. Afterwards, the face model is adapted by

$$P''''^{(i)} = P'''^{(i)} + \sum_i AUi(k_i) . \tag{15}$$

4. EXPERIMENTAL RESULTS

For testing the performance of the proposed algorithm, it has been applied to the three typical videophone sequences *Akiyo, Claire* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10Hz. Some results of these three sequences obtained by the proposed algorithm are shown in Fig. 4. Fig. 4a shows the original face, Fig. 4b the estimated eye and mouth features according to [14][15][16][17] and Fig. 4c the adapted face model *Candide* set onto the face. It can be seen from Fig. 4c that the adapted face model matches the estimated eye and mouth features very well.



(a) (b) (c)

Fig. 4 Some results of the three sequences obtained with the proposed algorithm. (a) Original face; (b) Estimated eye and mouth features; (c) Adapted face model *Candide* on the original face.

5. CONCLUSIONS

In this contribution, an algorithm for automatic adaptation of the face model *Candide* to the eye and mouth features using action units for semantic coding has been presented. This algorithm consists of three steps, namely: (i) the global adaptation; (ii) the local adaptation; (iii) the mimic adaptation. The global adaptation estimates the 3D eye and mouth center positions for scaling, rotation and translation of the face model in the 3D model world. In the second step, the eye corner points and the lip thickness of the face model are adapted. The eyes are locally scaled and inclined according to the estimated eye corner positions. Control vertices of the lip are shifted according to the estimated lip thickness. The final step is the mimic adaptation, in which the opening heights of the eyes and the mouth as well as the mouth corner positions of the face model are adapted by means of six action units. For this, the intensity parameters for each action unit are estimated based on the estimated eye and mouth features.

In contrast to [5][7][8][11], the global adaptation allows the face model to be rotated in the 3D model world, while no rotation of the face model is allowed in [8][11] and the rotation of the face model around x–axis is not allowed in [5][7]. Furthermore, the refined fit is arrived by the local and the mimic adaptation in this contribution. This can be used as basis for the mimic analysis by means of action units in the subsequent frames compared to [5][7][8].

This adaptation algorithm makes the implementation of a semantic coder realizable. In combination with a mimic analysis in the subsequent frames, a complete semantic coder shall be built for which a higher coding efficiency than for the coder without mimic analysis is aspired. These topics will be investigated in the future.

REFENRENCES

[1] K. Aizawa, H. Harashima and T. Saito, "Model–based analysis synthesis image coding (MBASIC) system for a person's face", *Signal Processing: Image Communication*, Vol. 1, No. 2, October 1989, pp.139–152.

[2] R. Chellappa, C.L. Wilson, S. Sirohey, "Human and machine recognition of faces: A survey", *Proceedings of IEEE*, Vol. 83, No. 5, May 1995, pp. 705–740.

[3] P. Eckman, V.W. Friesen, Facial coding system, Consulting Psychologists Press Inc., Palo Alto, U.S.A., 1977.

[4] J. Fischl, B. Miller, J. Robinson, "Parameter tracking in a muscle–based analysis/synthesis coding system", *Picture Coding Symposium* (PCS '93), Lausanne, Switzerland, No. 2.3, March 1993.

[5] M. Kampmann, R. Farhoud, "Precise face model adaptation for semantic coding of videophone sequences", *Picture Coding Symposium* (PCS '97), Berlin, Germany, Sept. 1997.

[6] 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", *Signal Processing: Image Communication*, Vol. 9, No. 3, March 1997, pp. 201–220.

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

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

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

[10] A. Samal, P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey", *Pattern Recognition*, Vol. 25, No. 1, January 1992, pp. 65–77.

[11] L. Strub, J. Robinson, "Automated facial conformation for model–based videophone coding", *IEEE International Conference on Image Processing*, Washington, DC, Vol. II, October 1995, pp. 587–590.

[12] W.J. Welsh, S. Searby and J.B. Waite, "Model–based image coding", *British Telecom Technol. J.*, Vol. 8, No. 3, July 1990, pp. 94–106.

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

[14] L. Zhang, "Estimation of eye and mouth corner point positions in a knowledge–based coding system", *Digital Compression Technologies and Systems for Video Communications,* Berlin, SPIE Vol. 2952, October 1996, pp. 21–28.

[15] L. Zhang, "Tracking a face for knowledge–based coding of videophone sequences", *Signal Processing: Image Communication*, Vol. 10, Nos. 1–3, July 1997, pp. 93–114.

[16] L. Zhang, "Estimation of the mouth features using deformable templates", *1997 International Conference on Image Processing*, Oct. 1997 Santa Barbara, CA.

[17] L. Zhang, "Automatic estimation of eye and mouth features for adaptation of a face model", submitted to *IEEE Transactions on Image Processing*.