Precise Face Model Adaptation for Semantic Coding of Videophone Sequences

Markus Kampmann, Ridha Farhoud
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
email: kampmann@tnt.uni–hannover.de, WWW: http://www.tnt.uni–hannover.de/~kampmann

ABSTRACT: *In this contribution, an algorithm for the automatic adaptation of a 3D face model for semantic coding of videophone sequences at very low bit rates is presented. After automatic estimation of facial features from an image sequence, the face model is adapted to the eyes, mouth, eyebrows, nose and to the chin and cheek contours of the person's face in the sequence. Applying the presented algorithm to the videophone sequences "Akiyo" and "Miss America", the face model is successfully adapted from the eleventh and second frame on, respectively.*

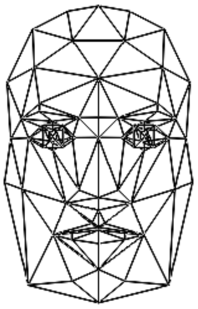
1. INTRODUCTION

For coding of video sequences at very low bit rates, an object–based analysis–synthesis coder (OBASC) has been introduced [1]. In an OBASC, real objects are described by model objects. A model object is defined by motion, shape and color parameters. The shape of a model object is represented by a 3D wireframe. The motion parameters describe translation and rotation of the model object in 3D space. The color parameters denote luminance and chrominance reflectance on the model object surface. These parameters are estimated automatically by image analysis [2]. No a–priori knowledge about the image content is exploited.

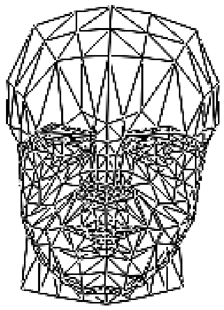
In typical videophone sequences, head and shoulders of human persons appear in the scene. This a–priori knowledge can be exploited in order to improve the coding efficiency. Therefore, an OBASC is extended in [3] to a knowledge–based analysis–synthesis coder (KBASC) by introduction of an automatic adaptation of the 3D face model *Candide* [4] (Fig. 1 (a)) to a person's face in the scene. At the beginning of the image sequence, the positions of the eyes and mouth centers are estimated. For adaptation, the face model is scaled and inclined using the estimated center positions and incorporated into the 3D wireframe. In order to achieve a higher coding efficiency, it is planned to extend the knowledge–based analysis–synthesis coder to a semantic coder [5]. In a semantic coder, facial expressions of a person are described by mimic parameters e.g. action units [6] or facial muscle parameters [7]. These parameters have to be estimated, coded and then transmitted to the receiver. At the receiver, the facial expressions are synthesized using the decoded parameters. The estimation of mimic parameters requires an automatic, very accurate adaptation of the face model to the person's face. Therefore, the face model has to be adapted not only to the eyes and mouth centers but also to the indivual shape of the person's face, e.g. to the face outline and to the nose. Furthermore, for initialization of the mimic analysis the face model should be adapted to the person's mimic, e.g. mimic of eyes, mouth and eyebrows.

In the literature, some algorithms are proposed for precise face model adaptation. In [7][8], face models are adapted manually. In [9], a face model is adapted by matching a texture template of an average face to the person's face in the scene. Rotations of the head are not considered. The algorithm in [10] adapts *Candide* to the eyes, mouth and chin contour of the individual face. This algorithm assumes that the person looks straight into the camera. Furthermore, the mouth of the person is assumed to be closed. In [11], the face model *Candide* is adapted to the eyes and mouth of the person's face by using action units.

In this contribution, an automatic algorithm for the precise adaptation of a 3D face model to an individual face is presented. The face model from [7] is used (Fig. 1 (b)), which incorporates a model of the human facial muscles for the synthesis of facial expressions. The face model is adapted to the eyes, mouth, chin and cheek contours, eyebrows and to the nose of a face in an image sequence. The automatic estimation of these facial features from an image has been carried out in previous work [3][12][13][14] and is not the subject of this contribution. The adaptation will not assume a person looking straight into the camera. For adaptation, a global adaptation will be carried out first. Hereby, the size of the face model and its orientation in 3D space shall be adapted. After global adaptation, a local adaptation will be adapt the face model to the estimated facial features.



(a)



(b)

Fig. 1: Face models: (a) *Candide* [4], (b) face model based on [7].

2. GLOBAL ADAPTATION

The global adaptation consists of three processing steps. In the first step, the face model is rotated around the y–head axis (see Fig. 2). Therefore, the eyes center positions $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ and the cheek contours in the image plane are evaluated. By a rotation around the y–axis, a deviation between $\boldsymbol{p}_3$, the middle between the two eyes centers, and $\boldsymbol{p}_6$, the middle between the face contour points $\boldsymbol{p}_4$ and $\boldsymbol{p}_5$, occurs (see Fig. 3). If the face is looking straight into the camera, $\boldsymbol{p}_3$ corresponds with $\boldsymbol{p}_6$. Therefore, a measure for the amount of rotation

$$\alpha_{real} = \frac{|\boldsymbol{p}_3 - \boldsymbol{p}_6|}{|\boldsymbol{p}_4 - \boldsymbol{p}_5|} \qquad (1)$$

of the real face is introduced. By projection of the face model onto the image plane, a similar measure for the amount of rotation

$$\alpha_{model} = \frac{|\boldsymbol{p}'_3 - \boldsymbol{p}'_6|}{|\boldsymbol{p}'_4 - \boldsymbol{p}'_5|} \qquad (2)$$

of the face model can be calculated (see Fig. 4). Now, the face model is iteratively rotated around the y–axis until

$$|\alpha_{model} - \alpha_{real}| < th_{\alpha} \qquad (3)$$

with a given threshold th_{α}.

In the second step, the face model is rotated around the z–head axis with a rotation angle β such that the lines through the estimated eyes center positions and through the projected positions of the eyes of the face model are parallel (see Fig. 3).

In the third step, the face model is scaled and afterwards shifted that the projection of the eyes and mouth centers of the face model corresponds with the estimated eyes and mouth center positions in the image plane.

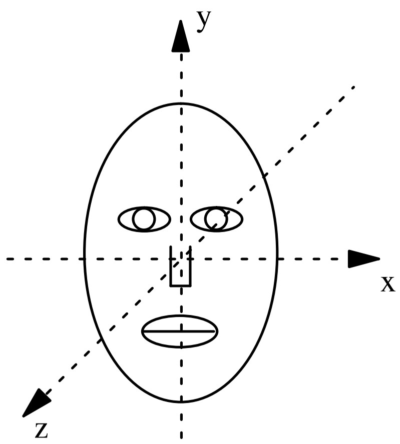


Fig. 2: Head rotation axes.

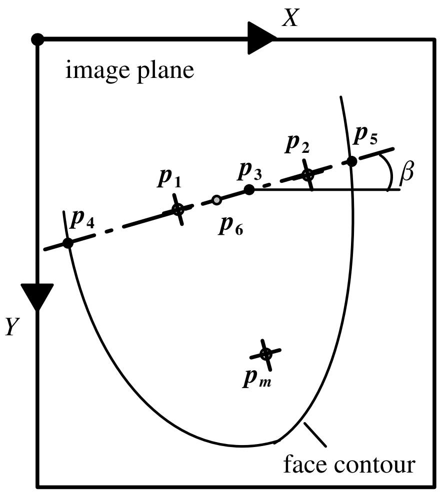


Fig. 3: Facial features in the image plane:
$\boldsymbol{p}_1, \boldsymbol{p}_2$: estimated eyes center positions,
$\boldsymbol{p}_3$: middle between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$,
$\boldsymbol{p}_4, \boldsymbol{p}_5$: estimated face contour points,
$\boldsymbol{p}_6$: middle between $\boldsymbol{p}_4$ and $\boldsymbol{p}_5$,
$\boldsymbol{p}_m$: estimated mouth center position.

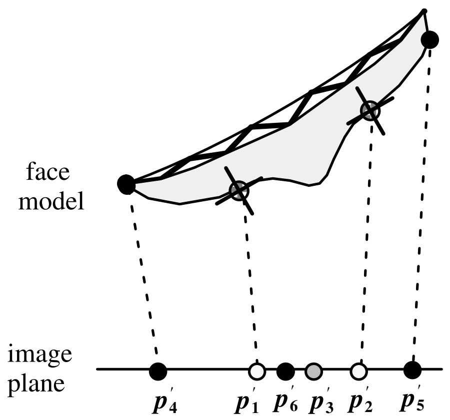


Fig. 4: Projection of the face model onto the image plane:
p_1', p_2' : eye center positions
p_3' : middle between p_1' and p_2'
p_4', p_5' : face contour points
p_6' : middle between p_4' and p_5'

3. LOCAL ADAPTATION

The local adaptation is carried out in the following steps.

3.1 Adaptation of the jaw

First, according to the estimated mouth's mimic the mouth of the face model is opened by a movement of the face model's jaw.

3.2 Adaptation of the face contour

In the next step, the face model is adapted to the estimated chin and cheek contours. Therefore, the control points of the chin and the cheek of the face model have to be shifted towards the estimated chin and cheek contours. Neighbouring control points of the face model have also to be shifted in order to keep the proportions of the human face. The control points of the eyes and mouth centers of the face model must not be shifted, because they have been already adapted by global adaptation.

For calculation of the 3D displacement vector of each control point, a 2D displacement vector in the image plane is calculated first. Then, the 3D displacement vector results from the 2D displacement vector by perspective projection.

For calculation of the 2D displacement vector of each control point, the face model is projected onto the image plane and so called reference displacement vectors (RDV) are determined (Fig. 5). A RDV is a 2D displacement vector in the image plane attached to each pel $\boldsymbol{r}_i$ of the projected face model contour C_{model}. It describes the smallest distance $\boldsymbol{v}_i$ between $\boldsymbol{r}_i$ and the estimated face contour C_{real}. Furthermore, three RDVs with $\boldsymbol{v}_i \equiv \boldsymbol{0}$ are attached to the eyes and mouth centers, since these centers have been already adapted by global adaptation. The 2D displacement vector $\boldsymbol{d}_j$ of each control point $\boldsymbol{c}_j$ is now interpolated by

$$\boldsymbol{d}_j = \frac{\sum_{i=0}^{I} \frac{1}{s_i} \boldsymbol{v}_i}{\sum_{i=0}^{I} \frac{1}{s_i}} \tag{4}$$

with s_i as the distance between $\boldsymbol{r}_i$ and the control point $\boldsymbol{c}_j$ in the image plane and I as the number of the RDVs (Fig. 6). Finally, the 3D displacement vector of each control point is calculated from $\boldsymbol{d}_j$ taking into account perspective projection.

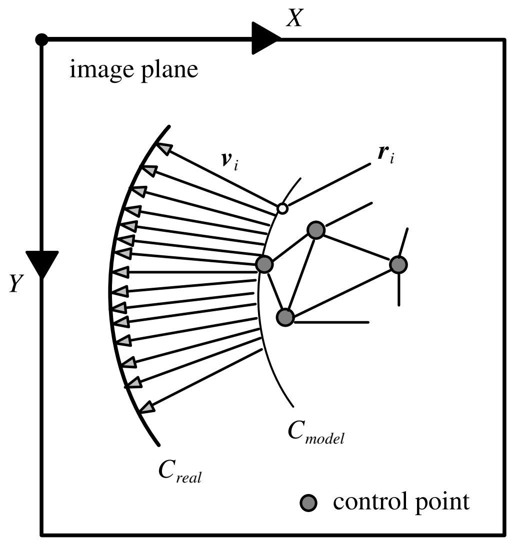


Fig. 5: Reference displacement vectors (RDVs) (C_{real}: estimated contour of the real face, C_{model}: contour of the projected face model).

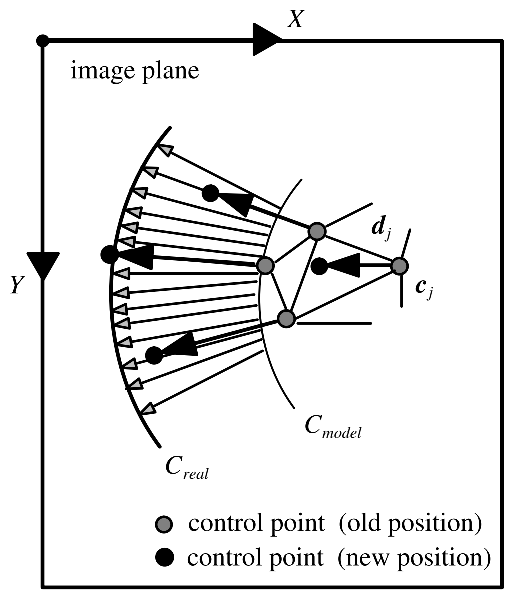


Fig. 6: 2D displacement vector $\boldsymbol{d}_j$ for a control point $\boldsymbol{c}_j$ (C_{real}: estimated contour of the real face, C_{model}: contour of the projected face model).

3.3 Adaptation of eyes, mouth, eyebrows and nose

In the next step, the eye lids of the face model are moved towards the estimated eye lid contours. Then, the control points of the lips are shifted towards the estimated mouth lip contours. Finally, the face model is adapted to the estimated sides of the nose and to the estimated eyebrows borders by moving the corresponding control points.

4. EXPERIMENTAL RESULTS

The described algorithm has been applied to the test sequences *Akiyo* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10Hz. For the test sequence *Akiyo* the facial features are estimated after 11 frames (Fig. 7(a)(b)). Afterwards, the face model is adapted to the estimated features applying the proposed algorithm. Fig. 7(c)(d)(e) shows the adapted face model. For the test sequence *Miss America* the facial features are estimated after 2 frames (Fig. 8(a)(b)). Afterwards, the face model is adapted (Fig. 8 (c)(d)(e)).

5. CONCLUSIONS

An algorithm for automatic adaptation of a face model to a person's face in a videophone sequence has been presented. The face model is adapted to the eyes, mouth, eyebrows, nose and chin and cheek contours of the individual face in the sequence. The adaptation algorithm consists of two processing steps. In the first step, a global adaptation is carried out adapting the size of the face model and its orientation in 3D space. In the second step, the face model is adapted to the different facial features by local adaptation. Applying the described adaptation algorithm to the test sequences *Akiyo* and *Miss America* (CIF,10Hz), the face model is adapted from the eleventh and second frame on, respectively. In the future, an algorithm for the analysis of facial expressions will be developed in order to complete a semantic coder.

6. REFERENCES

[1] H.G. Musmann, M. Hötter, J. Ostermann, "Object–oriented analysis–synthesis coding of moving images", *Signal Processing: Image Communications*, Vol. 3, No. 2, November 1989, pp. 117–138.

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

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

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

[5] H. 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.

[6] P. Ekman, W. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto CA., 1977.

[7] 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.

[8] 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.

[9] 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.

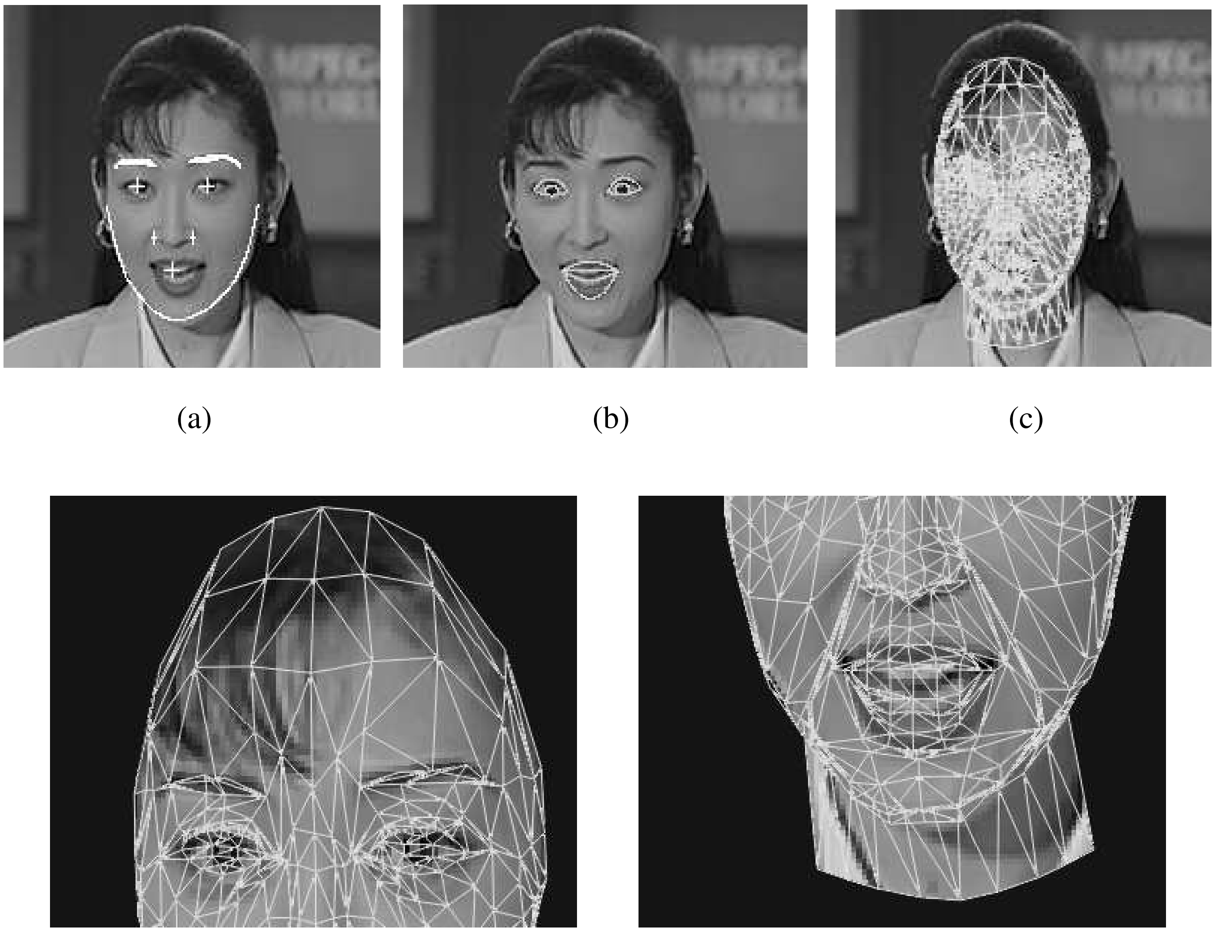
[10] M.J.T. Reinders, P.J.L. van Beek, B. Sankur, J.C. van der Lubbe, "Facial feature location and adaptation of a generic face model for model–based coding", *Signal Processing: Image Communication*, Vol. 7, No. 1, March 1995, pp. 57–74.

[11] L. Zhang, "Automatic adaptation of a face model using action units", *Picture Coding Symposium (PCS'97),* Berlin, Germany, Sept. 1997.

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

[13] M. Kampmann, "Estimation of the chin and cheek contours for precise face model adaptation ", *International Conference on Image Processing (ICIP97)*, Santa Barbara, USA, Oct. 1997.

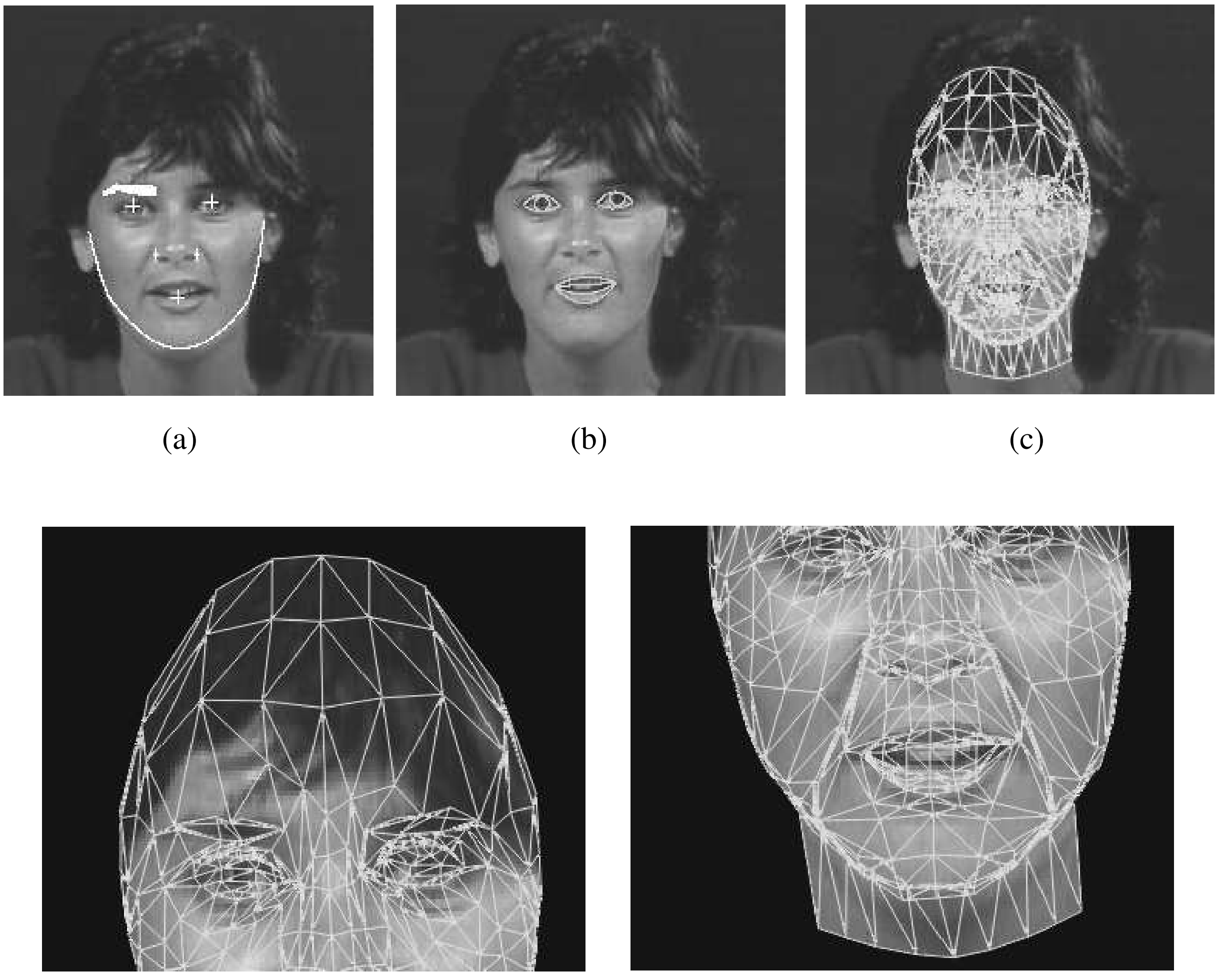
[14] V. Chanawatr, "Modellierung und Segmentierung von Nase und Augenbrauen in Bildtelefonsequenzen", Studienarbeit, University of Hannover, June 1997.



(a) (b) (c)

(d) (e)

Fig. 7: Test sequence *Akiyo* (CIF, 10Hz, 11th frame): (a) estimated eyes and mouth centers, chin and cheek contours, eyebrows and sides of the nose, (b) estimated eye lids and mouth lips, (c) adapted face model, (d) top part of the adapted face model, (e) bottom part of the adapted face model.



(a) (b) (c)

(d) (e)

Fig. 8: Test sequence *Miss America* (CIF, 10Hz, 2nd frame): (a) estimated eyes and mouth centers, chin and cheek contours, eyebrows and sides of the nose, (b) estimated eye lids and mouth lips, (c) adapted face model, (d) top part of the adapted face model, (e) bottom part of the adapted face model.