VLBV 97
Linköping , Sweden, July 1997

Automatic Adaptation of face models in videophone sequences with more than one person

Markus Kampmann, Geovanni Martinez
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
Phone: ++49–511–7625314, Email: kampmann@tnt.uni–hannover.de
WWW: http://www.tnt.uni–hannover.de/~kampmann

Summary

For coding of moving images at 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. These parameters are estimated automatically by image analysis. By the source model of moving 3D objects[2], 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 of the model object surface. Moreover, 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 knowledge can be exploited in order to improve the modelling accuracy for this kind of scenes. Therefore, OBASC is extended in [3] to a knowledge–based analysis–synthesis coder (KBASC) by adaptation of the 3D face model *Candide* [4] to a person in the scene. In order to adapt *Candide,* the positions of eyes and mouth have to be estimated. First, assuming that only one person appears in the scene, the head area is extracted by evaluating the silhouette of the person and assuming a wide upper part of the body and a narrower head. Then, the eyes and mouth positions are estimated using template matching and feature extraction techniques. Finally, the face model is adapted and incorporated into the 3D model object. This algorithm fails, when more than one person appears in the scene. Other approaches for the estimation of eyes and mouth positions in videophone sequences are restricted to one person too [5][6][7][8].

In this contribution, the algorithm in [3] is extended to allow the automatic adaptation of face models in videophone sequences with more than one person in the scene. Moreover, the persons could be placed close together and could share a common silhouette. In order to adapt *Candide,* the eyes and mouth positions of several persons in the image sequence have to be estimated. The new adaptation algorithm consists of five processing steps. In the first step, human heads are detected. Therefore, the 3D model objects are subdivided by object articulation into 3D object components (Fig. 1 (a)(b)). For object articulation, neighbouring triangles which exhibit similar 3D motion during the image sequence are clustered into object components [9][10]. This is carried out without using a–priori knowledge about the image content. Afterwards, it is checked whether an object component could be a human head. In order to be recognized as a human head, an object component must fulfill several conditions. The silhouette of the object component has to be roughly a circle. Furthermore, a human head must be located above another object component, i. e. the upper part of the human body. After detection of human heads, the following processing steps are carried out for each head. In the second step, the mouth center position is estimated by evaluating horizontal contours in the lower part of the head area. In the third step, the pupils as the eyes center positions are estimated in the upper part of the head area. First, areas with horizontal contours are extracted. In these areas, a template matching with a luminance template of an average eye is carried out. The pupils represent the darkest image points in regions with high correlation with the eye template. In the forth step, the estimated eyes and mouth center positions are verified. They have to span an isosceles triangle in order to be accepted as correct eyes and mouth center positions. Otherwise, these positions are rejected and an adaptation of the face model is not possible for this frame. If the verification of the centers of eyes and mouth is successfully, the face model *Candide* is finally adapted in the fifth step. According to the estimated eyes and mouth center positions, the face model is scaled and inclined. Then, it is incorporated into the 3D model object.

The described algorithm has been applied to the test sequences *Akiyo*, *Claire* and to the test sequence with two persons *Anis_Markus* with a spatial resolution corresponding to CIF and a frame rate of 10Hz (Fig. 1, 2, 3). Applying the proposed algorithm to the test sequence *Anis_Markus*, both human heads are detected after 10 frames (Fig. 1 (a)(b)). For the right person, the face model *Candide* is adapted immediately (Fig. 1 (c)(d)), for the left person at the 12th frame (Fig. 1 (e)(f)). Using the algorithm in [3], an adaptation of face models for *Anis_Markus* is not possible. For the test sequence *Claire*, the face model is adapted after 6 frames using the proposed algorithm (Fig. 2 (b)(c)). Applying the algorithm in [3], 10 frames are required. For the test sequence *Akiyo*, both algorithms adapt the face model after 7 frames (Fig. 3 (b)(c)).

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. Huang, M. Ouhyoung, J. Wu,”Automatic feature point extraction on a human face in model–based image coding”, Optical Engineering Journal, Vol. 32, No. 7, July 1993, pp. 1571–1580.

[6] P.J.L. van Beek, M.J.T. Reinders, B. Sankur, J.C.A. van der Lubbe,”Semantic Segmentation of videophone image sequences”, SPIE Vol. 1818, *Visual Communications and Image Processing’92*, Boston, USA, November 1992, pp. 1182–1193.

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

[8] F. Mu, H. Li, R. Forchheimer,”Automatic extraction of human facial features”, *Signal Processing: Image Communications*, Vol.8, No. 4, May 1996, pp. 309–326.

[9] G. Martínez, ”Shape Estimation of Articulated 3D Objects for Object–Based Analysis–Synthesis Coding (OBASC)”, *Signal Processing: Image Communications*, Vol. 9, No. 3, March 1997, pp. 175–199.

[10] G. Martínez, ”Shape Estimation of Articulated 3D Objects considering mutual occlusions for object–based analysis–synthesis coding (OBASC)”, *Picture Coding Symposium 96 (PCS’96),* Melbourne, Australia, March 1996, pp. 141–146.

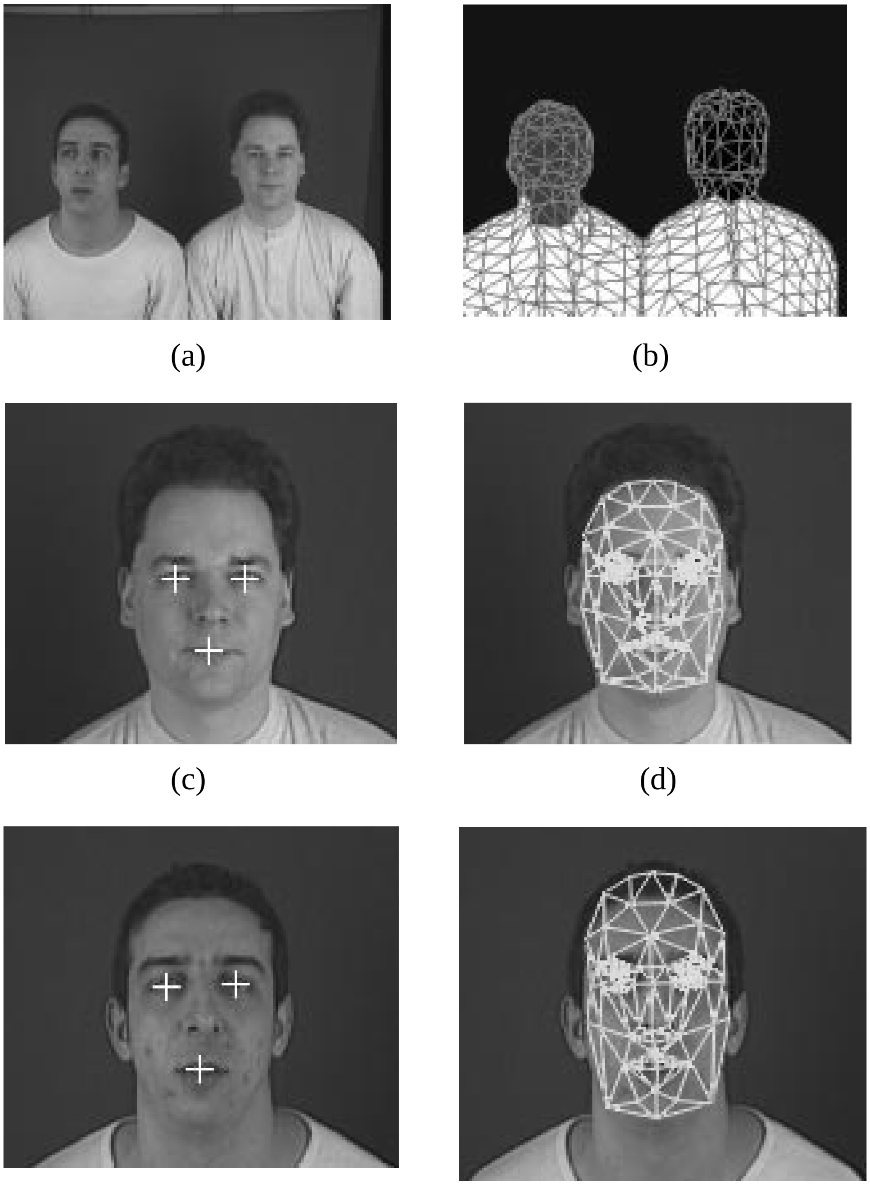

Fig. 1: Testsequence *Anis_Markus* (CIF, 10Hz): (a) 10th frame, (b) subdivision of the 3D model object into 3D object components (10th frame), (c) estimated eyes and mouth positions for the right person (10th frame), (d) adapted face model for the right person (10th frame), (e) estimated eyes and mouth positions for the left person (12th frame), (f) adapted face model for the left person (12th frame).

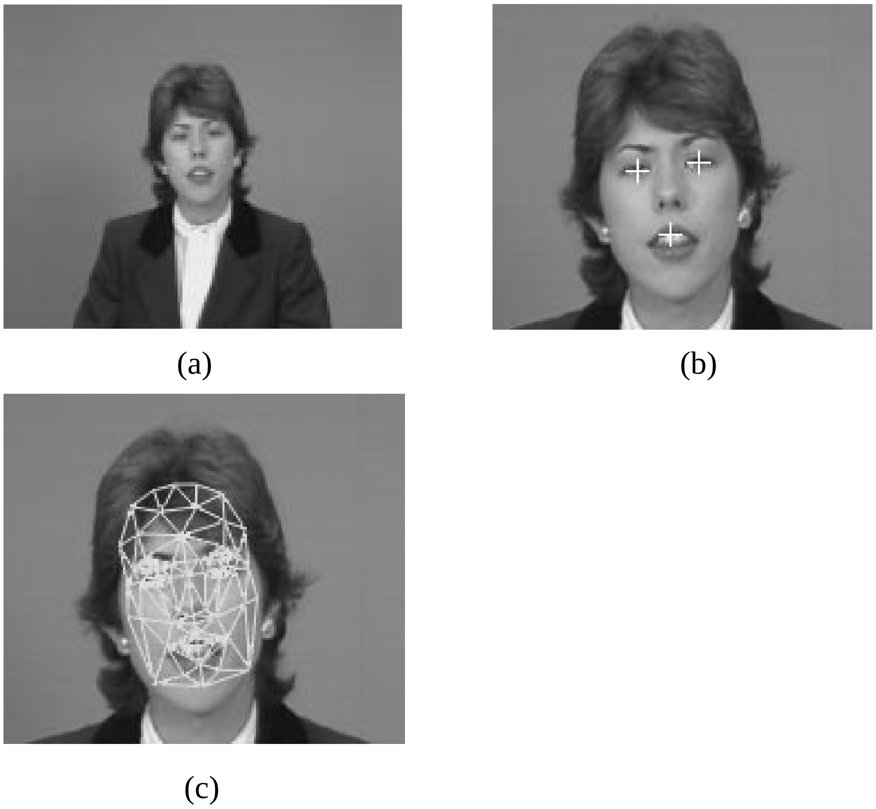

Fig. 2: Test sequence *Claire* (CIF, 10Hz, 6th frame): (a) original image, (b) estimated eyes and mouth positions, (c) adapted face model.

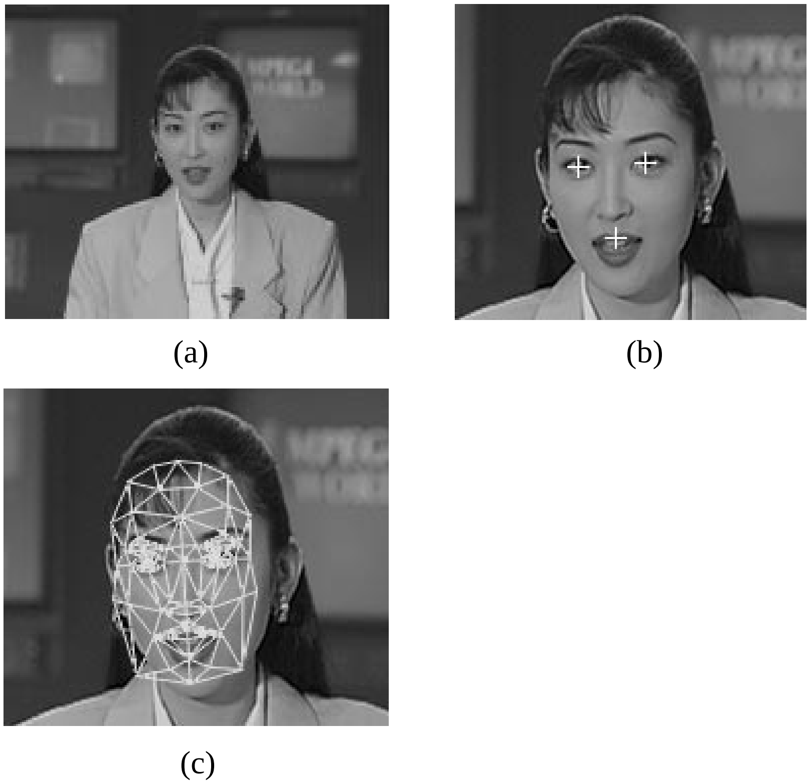

Fig. 3: Test sequence *Akiyo* (CIF, 10Hz, 7th frame): (a) original image, (b) estimated eyes and mouth positions, (c) adapted face model.