Estimation of the Chin and Cheek Contours for Precise Face Model Adaptation

Markus Kampmann
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

For semantic coding of videophone sequences at very low bit rates, the adaptation of a face model to automatically estimated facial features is necessary. In this contribution, an algorithm for the automatic estimation of the chin and cheek contours of a person is presented. The chin and cheeks are represented by a deformable template consisting of parabolas. Costfunctions are established and minimized to find the best fit of the template to the chin and cheek contours of the person. In all examined frames of the videophone sequences "Akiyo" and "Claire", the manually determined real contour of the chin and the cheeks is approximated by the estimated contours with good accuracy. In more than 60% of all frames, no deviation between the real contours and the estimated contours occurs.

1. Introduction

For coding of moving images 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, which are estimated automatically. 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, the color parameters denote luminance and chrominance reflectance of the model object surface.

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* to a person's face in the scene. At the beginning of the image sequence, the positions of eyes and mouth of the person in the scene are estimated. Afterwards, the face model is adapted to the person's face applying the estimated positions of eyes and mouth. After adaption, the wireframe with the face model is motion compensated throughout the image sequence. For an automatic analysis of the person's facial expressions, a more accurate adaptation of the face model to the chin and cheek contours of the individual person is necessary. Therefore, the chin and cheek contours in the image sequence have to be estimated.

For estimation of chin and cheek contours, several algorithms have been proposed. In [4][5], active contour models (*snakes*) are used for the estimation of the chin and cheek contours. A *snake* is an energy–minimizing spline influenced by image features to pull it toward edges. These approaches were applied to persons looking straight into the camera. The reliability of these algorithms are low [5]. In [6], only the chin contour is estimated by using the concept of deformable templates [7]. In [6], the deformable template for the chin consists of two parabolas. A cost function is minimized to find the best fit of the template to the chin. For this cost function it is assumed that the person in the scene is looking straight into the camera.

In this contribution, an algorithm for the estimation of the chin and cheek contours of persons in videophone sequences is proposed which is not limited to persons looking straight into the camera. The algorithm evaluates eyes and mouth positions that are known from the tracked face model. The chin and the cheeks are represented by a deformable template consisting of four parabolas (Fig. 1). The two lower parabolas define the chin template and are described by six parameters, the position of the origin A and the positions of the endpoints B and C of the parabolas. The two upper parabolas define the left and the right cheek template. They are linked together with the chin template and are described by their endpoints D and E. The endpoints D and E are located both on the line l_4 (see Fig. 1). Line l_4 is a parallel to the line l_3 which connects the eyes centers. The distance g between l_3 and l_4 is calculated from the known eyes and mouth positions.

Section 2 describes the estimation of the chin contours, Section 3 the estimation of the cheek contours. Experimental results with typical videophone sequences are given in Section 4.

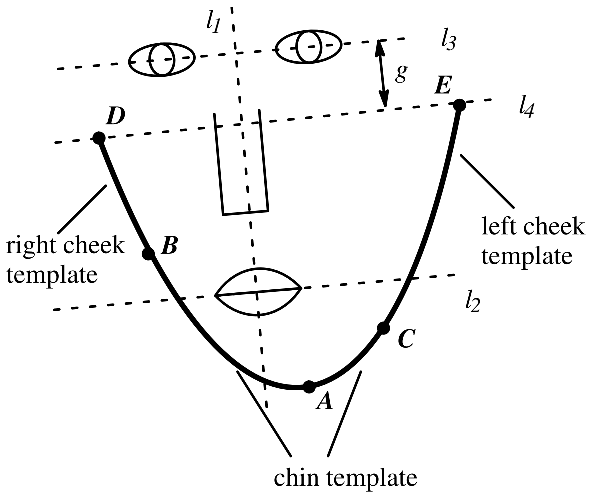


Fig. 1: Deformable template for the chin and cheek contours.

2. Estimation of chin contours

The algorithm for the estimation of the chin contours consists of three processing steps.

In the first step, an initial position for the chin template is estimated. In order to reduce the complexity of the parameter estimation problem some constraints are introduced. The origin A must be part of line l_1 through the middle between the eyes and through the mouth center (see Fig. 1). Furthermore, the endpoints B and C of the chin parabolas must be part of line l_2, where l_2 goes through the mouth center and is a parallel to the line l_3 (see Fig. 1). Taking into account the tracked eyes and mouth positions the probability of the occurence of the chin contour at a certain position in the image can be modeled empirically due to the anatomy of an average face. Assuming the anatomy of an average face, the origin A must be part of the line between a_1 and a_2 (Fig. 2). The probability of the occurence of A at a certain position on the line between a_1 and a_2 is modelled with the highest probability for A in the middle between a_1 and a_2. For the endpoints B and C, similar probability functions are established (Fig. 2). If a person is not looking straight into the camera, the distances between the mouth center and the left and right chin contour become unequal (Fig. 3). However, the distance between the left and right chin contour is approximately constant (Fig. 3). Therefore, an additional probability function for the endpoints B and C of the chin template is introduced. The distance between B and C has a minimal value h_1 and a maximum value h_2 (Fig. 4). The probability of a certain value of the distance between B and C is modelled with the highest probability at a distance $(h_1+h_2)/2$.

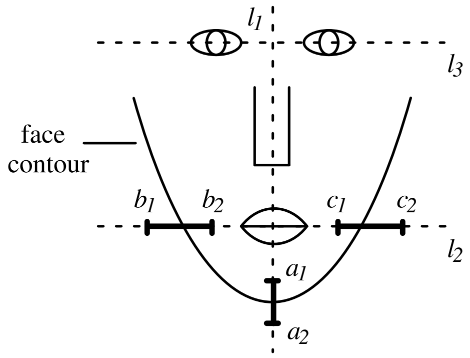


Fig. 2: Estimation of the chin contours (step 1): A must be part of line a_1a_2, B of line b_1b_2, C of line c_1c_2.

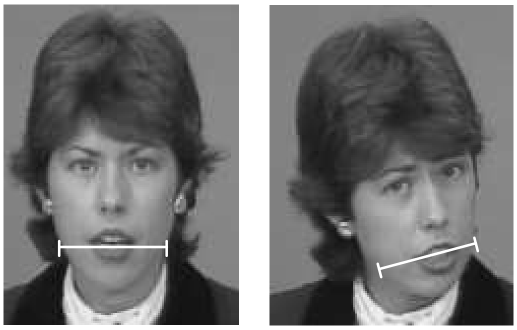


Fig. 3: Estimation of chin contours (step 1): distance between left and right chin contour is approximately constant despite head rotation.

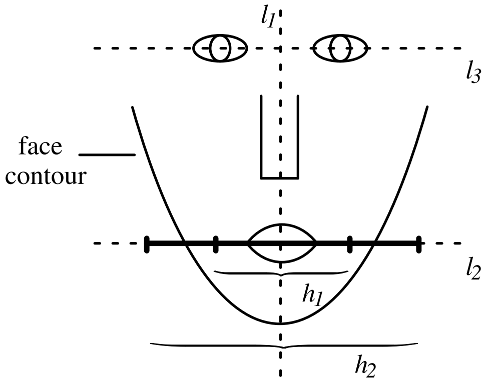
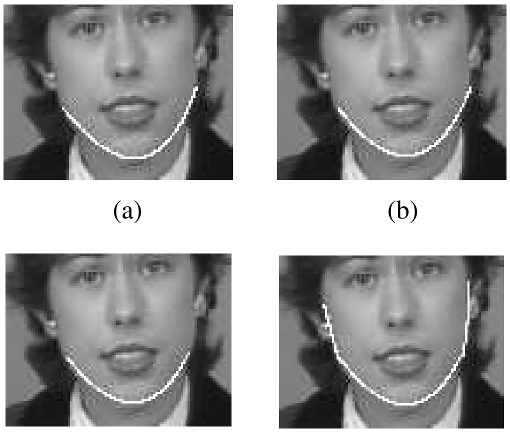


Fig. 4: Estimation of the chin contour (step 1): minimum distance h_1 and maximum distance h_2 between the endpoints B and C of the chin template.

A cost function f_1 is established using these probabilities and additionally preferring high values of the image gradient at the position of the chin contour. By minimizing this cost function, the initial position of the chin template is estimated (Fig. 5 (a)).



(a) (b) (c) (d)

Fig. 5: Estimated chin contours after (a) step 1, (b) step 2, (c) step 3; (d) estimated chin and cheek contours.

In the second step, the final value for the origin A is estimated. The assumption that A is part of l_1 is omitted now. The origin A is estimated by minimizing a further costfunction f_2. Additionally to the costfunction f_1 in the first step, the costfunction f_2 prefers a small distance between the initial position of A after the first step and the final position of A (Fig. 5 (b)).

In the third step, the final values for the endpoints B and C are estimated. The assumption that B and C are part of l_2 is also omitted now. Additionally to the terms in the previous cost functions, the cost function f_3 in this step takes into account the average luminance value of the face (Fig. 5 (c)).

3. Estimation of cheek contours

After estimation of the chin contours, the parameters for the left and the right cheek template are estimated. Similar to the estimation of the parameters of the chin template, the endpoints D and E must be part of the lines between d_1 and d_2 and e_1 and e_2, respectively (Fig. 6). Furthermore, the distance between D and E is approximately constant despite head rotation. Assuming the anatomy of an average face, probabilities for the occurence of D and E at a certain position are modelled and a costfunction f_4 is established using these probabilities and additionally preferring high values of the image gradient at the position of the cheek contour. Finally, the parameters for the cheek templates are estimated by minimizing the costfunction f_4 (Fig. 5 (d)).

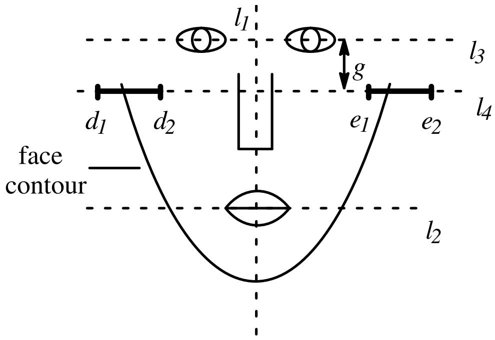
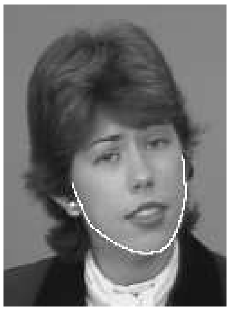


Fig. 6: Estimation of the cheek contours: D must be part of $d_1 d_2$, E of $e_1 e_2$.

4. Experimental results

The described algorithm has been applied to the first 50 frames of the test sequences *Akiyo* and *Claire* with a spatial resolution corresponding to CIF and a frame rate of 10Hz. Fig. 7 shows representative results of the estimated chin and cheek contours. In all examined frames the manually determined real contour of the chin and the cheeks is approximated by the estimated contours with good accuracy despite that the persons do not always look straight into the camera. In more than 60% of all examined frames, no deviation between the manually determined contours and the estimated contours occurs. In the remaining frames, only small estimation errors are noticed. Fig. 7 (b) and Fig. 7 (f) show the results with the maximum estimation error for all examined frames.

(a)



(d)

(b)

(e)

(c)

(f)

Fig. 7: Estimated chin and cheek contours for the test sequence *Akiyo* (CIF, 10Hz)
(a) 14th frame, (b) 39th frame, (c) 50th frame and the test sequence *Claire* (CIF, 10Hz)
(d) 31th frame, (e) 35th frame, (c) 50th frame.

5. Conclusions

An algorithm for the estimation of the chin and cheek contours of persons in videophone sequences has been presented which evaluates eyes and mouth positions that are known from an automatically tracked face model. The chin and cheek contours are represented by a deformable template consisting of four parabolas. First, the parameters of the chin template are estimated. Taking into account the tracked eyes and mouth positions the probability of the occurence of the chin contour at a certain position in the image can be modeled empirically due to the anatomy of an average face. Costfunctions take into account these probabilities and additionally preferring high values of the image gradient at the position of the chin contour. By minimizing these cost functions the parameters of the chin template are estimated. Similar to the estimation of the parameters of the chin template, the parameters of the cheek templates are estimated in the second step. In all examined frames of the videophone sequences *Akiyo* and *Claire* (CIF,10Hz), the manually determined real contour of the chin and the cheeks is approximated by the estimated contours with good accuracy. In more than 60% of all frames, no deviation between the real contours and the estimated contours occurs.

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.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.

[5] Chung–Lin Huang, Ching–Wen Chen, "Human facial feature extraction for face interpretation and recognition", *Pattern Recognition*, Vol. 25, No. 12, 1992, pp. 1435–1444.

[6] M.J.T. Reinders, F.A. Odijk, J.C.A. van der Lubbe, J.J. Gerbrands, "Tracking of global motion and facial expressions of a human face in image sequences", SPIE Vol. 2904, *Visual Communications and Image Processing'93*, Cambridge, MA, November 1993, pp. 1516–1527.

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