Estimation of the Mouth Features Using Deformable Templates

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

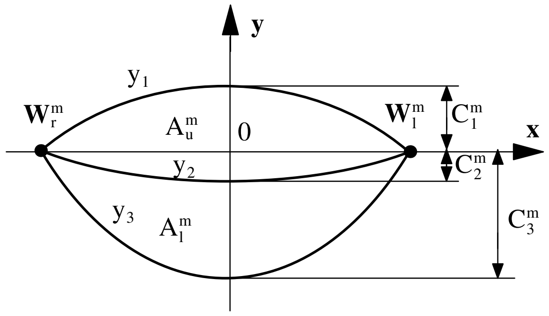
Abstract

Automatic estimation of mouth features is one of the important topics for face recognition and model–based coding of videophone sequences. In this contribution, an automatic mouth feature estimation algorithm which uses deformable templates is developed. Here, the mouth features are represented by the corner points of the mouth as well as the so called lip outline parameters. The lip outline parameters describe the opening of the mouth and the thickness of the lips. Compared to previous works, simplified cost functions are introduced. Furthermore, an algorithm for automatic determination of whether the mouth is open or closed is developed. Experimental results obtained with typical videophone sequence are given to evaluate the performance of the proposed algorithm. It is shown that in about 37% of the images the mouth features could be automatically estimated, which can be used e.g. for improving face modelling in a model–based coder.

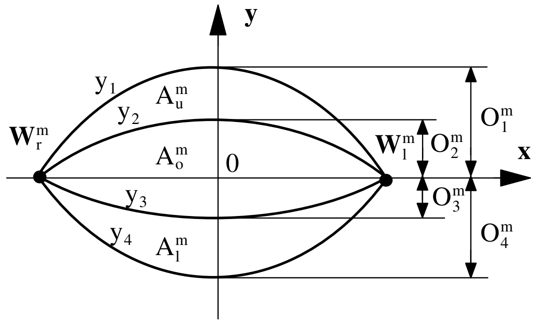
1. Introduction

Face recognition [2] and model–based coding of videophone sequences [1][3][4] require the automatic estimation of facial features. Several approaches to this problem have been published. Overviews can be found in [2][7]. For the estimation of mouth features, deformable template matching as proposed in [8] has been frequently applied [5][6]. A deformable mouth template is matched to the person's mouth by minimizing cost functions. There are two problems associated with this approach: (1) The cost functions to be minimized for finding a good match are complex, e.g. the mouth template matching for a closed mouth needs 7 weighting factors. The minimization is performed in several steps, and some of the weighting factors have different values in different steps[5]; (2) The selection of an appropriate mouth template, i.e. a closed or an open mouth template, is done manually.

In order to overcome these problems, a new algorithm for automatic estimation of the mouth features of a human face in a head–and–shoulder videophone sequence is proposed. The features of the mouth are described by a *mouth–closed* and a *mouth–open* deformable template consisting of



(a) *Mouth–closed* deformable template



(b) *Mouth–open* deformable template

Fig. 1 The parameterized deformable templates. W_l^m and W_r^m are the mouth corner points. C_i^m (i=1,2,3) and O_i^m (i=1,2,3,4) are the lip outline parameters.

three or four parabolas y_i, respectively (Fig. 1). The mouth features are represented by the mouth corner points and the so called lip outline parameters (C_i^m for a closed mouth and O_i^m for an open mouth). The lip outline parameters denote the opening of the mouth and the thickness of the lips.

In the proposed algorithm, these parameters are estimated in two steps. The first step estimates the corner points of the mouth as described in [9]. The second step, which is dealt with in this contribution, utilizes these estimated corner points of the mouth for performing deformable template matching, by which the lip outline parameters are estimated keeping the corner points fixed.

This contribution is organized as follows: after the introduction, the algorithm for estimating the lip outline parameters is described. Section 3 gives experimental results obtained with test sequence *Miss America* (CIF, 10 Hz). Conclusions can be found in Section 4.

2. Estimation of the lip outline parameters

The lip outline parameter estimator consists of four parts, namely: (i) the estimation of the candidates for the lip outline parameters, (ii) the determination of whether the mouth is open or closed, (iii) the deformable template matching, and (iv) the verification of the estimated lip outline parameters. In the next subsections, the elements of this estimator will be addressed. Here, the input image is supposed to be provided in a (Y, C_b, C_r) color space. In the following, m denotes mouth and l and u denote the lower and upper lip, respectively.

Candidates for the lip outline parameters

The estimation of the candidates for the lip outline parameters consists of two parts, namely the detection of the contours within the mouth region and the selection of the candidates for the lip outline parameters.

The contours within the mouth region are detected by first using a morphological edge detector, then binarizing and thinning. For this, only the *Y* component of the image is used.

For the selection of the candidates, a perpendicular bisection of the line l^m which connects the mouth corner points intersects the contours within the extracted mouth region at some points (Fig. 2). The intersection points which lie above the line l^m up to $0.3L^m$ are denoted as $\boldsymbol{P}_i^u$ $(i = 1, ..., N_u^m)$ and the intersection points which lie below the line l^m up to $0.5L^m$ as $\boldsymbol{P}_j^l$ $(j = 1, ..., N_l^m)$, whereby it is considered that the lower lip has larger movement than the upper lip and L^m is the width of the mouth. The distances from the intersection points $\boldsymbol{P}_i^u$ and $\boldsymbol{P}_j^l$ to the line l^m are taken as the candidates for the upper and lower lip outline parameters within the mouth region, respectively.

Determination of whether the mouth is closed or open

After the candidates for the lip outline parameters have been estimated, the lip outline parameters can be determined from these estimated candidates. Because there are two deformable mouth templates, an open mouth template and a closed mouth template, the situation whether the mouth is open or closed has to be determined first. If from

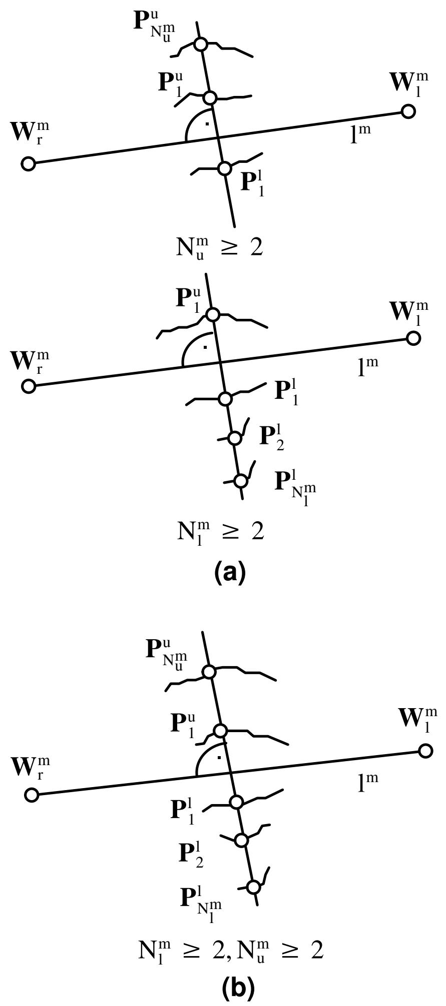


Fig. 2 Candidates for the lip outline parameters and determination of whether the mouth is closed or open, (a) the mouth is closed; (b) the mouth is open.

the candidates for the upper and lower lip outline parameters described above there is only one candidate for the lower lip outline parameter, i.e. $N_l^m = 1$, $N_u^m \geq 2$, or only one candidate for the upper lip outline parameter, i.e. $N_l^m \geq 2$, $N_u^m = 1$, the mouth is assumed to be closed (Fig. 2a). In case that there are at least two candidates for both the upper and the lower lip outline parameters, i.e. $N_l^m \geq 2$, $N_u^m \geq 2$, the mouth is assumed to be open (Fig. 2b). In other cases, the lip outline parameters can not be estimated.

Deformable template matching

After it has been determined whether the mouth is open or closed, deformable template matching can be performed to select the upper and lower lip outline parameters from the estimated candidates. For this purpose, different cost functions are proposed for an open and a closed mouth.

The mouth–open cost function

The cost function for an open mouth f_o^m is defined as:

$$f_o^m = (k_1 f_1 + k_2 f_2 + k_3 f_3) \rightarrow MIN \quad (1)$$

with f_1 being

$$f_1 = -\sum_{i=1}^{4} \frac{1}{L_{y_i}} \oint_{y_i} E_Y(X) ds \,, \quad (2)$$

where E_Y is the edge strength extracted from the Y component of the image using a morphological edge detector, L_{y_i} $(i = 1, 2, 3, 4)$ are the lengths of the parabolas (Fig. 1b), so that f_1 is the average of the edge strength over L_{y_i} $(i = 1, 2, 3, 4)$, f_2 being

$$f_2 = -\left|m_{A_u^m} - m_{A_o^m}\right| - \left|m_{A_l^m} - m_{A_o^m}\right| + \sigma_{A_u^m} + \sigma_{A_o^m} + \sigma_{A_l^m} \quad (3)$$

and f_3 being

$$f_3 = \left|\sigma_{A_u^m} - \sigma_{A_o^m}\right| + \left|\sigma_{A_o^m} - \sigma_{A_l^m}\right| + \left|\sigma_{A_l^m} - \sigma_{A_u^m}\right| , \quad (4)$$

where $m_{A_u^m}, m_{A_l^m}, m_{A_o^m}$ and $\sigma^2_{A_u^m}, \sigma^2_{A_l^m}, \sigma^2_{A_o^m}$ are the means and the variances of the C_r component of the image in the regions of the upper lip A_u^m, the lower lip A_l^m and in the region between the lips A_o^m (Fig. 1b), respectively. Since the teeth and the rest between the lips have very different values in the Y component but not in the C_r component, the C_r component is used here for evaluation rather than the Y component. Coefficents k_i $(i = 1, 2, 3)$ are the weighting factors which were set to 1 in the experiments. The term f_2 considers that the regions A_u^m, A_l^m and A_o^m have different C_r values but inside each region there are almost no differences of the C_r value. The term f_3 considers that the variance of the camera noise is the same in all regions. Now, the upper lip outline parameters (O_1^m, O_2^m) are selected from the candidates denoted $\boldsymbol{P}_i^u$ $(i = 1, ..., N_u^m)$ with $N_u^m \geq 2$ and the lower lip outline parameters (O_3^m, O_4^m) from the candidates denoted $\boldsymbol{P}_j^l$ $(j = 1, ..., N_l^m)$ with $N_l^m \geq 2$. For each possible combination $(O_1^m, O_2^m, O_3^m, O_4^m)$, f_o^m is calculated. That combination of $(O_1^m, O_2^m, O_3^m, O_4^m)$ with the minimum value of f_o^m is selected as the lip outline parameters of the open mouth.

The mouth–closed cost function

The cost function for a closed mouth is defined as

$$f_c^m = (k_1 f_1 + k_2 f_2 + k_3 f_3) \rightarrow MIN \quad (5)$$

with f_1 being

$$f_1 = -\sum_{i=1}^{3} \frac{1}{L_{y_i}} \oint_{y_i} E_Y(X) ds \,, \quad (6)$$

where L_{y_i} $(i = 1, 2, 3)$ are the lengths of the parabolas (Fig. 1a), f_2 being

$$f_2 = \left|m_{A_u^m} - m_{A_l^m}\right| + \sigma_{A_u^m} + \sigma_{A_l^m} \quad (7)$$

and f_3 being

$$f_3 = \left|\sigma_{A_u^m} - \sigma_{A_l^m}\right| . \quad (8)$$

Coefficients k_i $(i = 1, 2, 3)$ are the weighting factors which were set to 1 in the experiments. Lip outline parameters (C_1^m, C_2^m, C_3^m) are selected from the candidates $\boldsymbol{P}_i^u$ and $\boldsymbol{P}_j^l$. For each possible combination (C_1^m, C_2^m, C_3^m), f_c^m is calculated. That combination of (C_1^m, C_2^m, C_3^m) with the minimum value of f_c^m is selected as the lip outline parameters of the closed mouth.

Verification of the estimated lip outline parameters

After the lip outline parameters have been estimated, they have to be verified. Different criteria are exploited for an open or a closed mouth.

The mouth–open criterion

For an open mouth, the estimated lip outline parameters must satisfy

$$\left|O_1^m - O_2^m\right| \leq \left|O_4^m - O_3^m\right| \,, \quad (9)$$

$$m_{A_o^m} < m_{A_u^m} \quad , \quad m_{A_o^m} < m_{A_l^m} \quad , \quad (10)$$

where $m_{A_u^m}, m_{A_l^m}$ and $m_{A_o^m}$ are the means of the C_r component in the regions of the upper lip A_u^m, the lower lip A_l^m and in the region between the lips A_o^m (Fig. 1a), respectively. Ineq. (9) denotes that the upper lip can not be thicker than the lower lip, while Ineq. (10) stands for the fact that the red component within either lip is more intense than within the region between the lips.

The mouth–closed criterion

For a closed mouth, the estimated lip outline parameters must satisfy

$$\left|C_1^m - C_2^m\right| \leq \left|C_3^m - C_2^m\right| \,, \quad (11)$$

$$\left|m_{A_u^m} - m_{A_l^m}\right| \leq T \,, \quad (12)$$

where $m_{A_h^m}$ and $m_{A_l^m}$ are the means of the C_r component in the regions of the upper lip A_u^m and the lower lip A_l^m (Fig. 1a), respectively. In the experiments, threshold T is set to 1. Ineq. (11) denotes that the upper lip can not be thicker than the lower lip, while Ineq. (12) expresses that the two lips should be almost equally red. Ineq. (10) cannot be used for an open mouth because of the different reflections of the upper and the lower lip. If the estimated values do not pass the verification, they are rejected.

3. Experimental results

For testing the performance of the proposed algorithm, it has been applied to a videophone sequence *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10Hz. In about 37% of all frames (18 frames from 49 frames) the mouth features have been automatically estimated by the proposed algorithm. Fig. 3 shows the results obtained with the proposed algorithm. The decision between a closed and an open mouth and the estimation of the lip outline parameter is carried out automatically. In one frame from 18 frames an open mouth has been estimated closed (the last frame in the second line in Fig.3).

4. Conlusions

This contribution presents an algorithm for automatic estimation of mouth features. In contrast to [5][8], the cost functions have been simplified where the number of the weighting factors for the mouth templates has been reduced to 3. The values of all 3 weighting factors could simply be set to 1. Finally, geometrical criteria are exploited to verify the estimated lip outline parameters. It is shown that in a typical videophone sequence, in about 37% of the images the mouth features could be automatically estimated, which can be used e.g. for improving face modelling in a model–based coder[10].

References

[1] K. Aizawa, T.S. Huang, "Model–based image coding: Advanced video coding techniques for very low bit–rate applications", *Proceedings of the IEEE*, Vol. 83, No. 2, February 1995, pp. 259–271.

[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] H. Li, A. Lundmark, R. Forchheimer, "Image sequence coding at very low bitrates: A review", *IEEE Transactions on Image Processing*, Vol. 3, No. 5, September 1994, pp. 589–609.

[4] D.E. Pearson, "Developments in model–based video coding", *Proceedings of the IEEE*, Vol. 83, No. 6, June 1995, pp. 892–906.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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

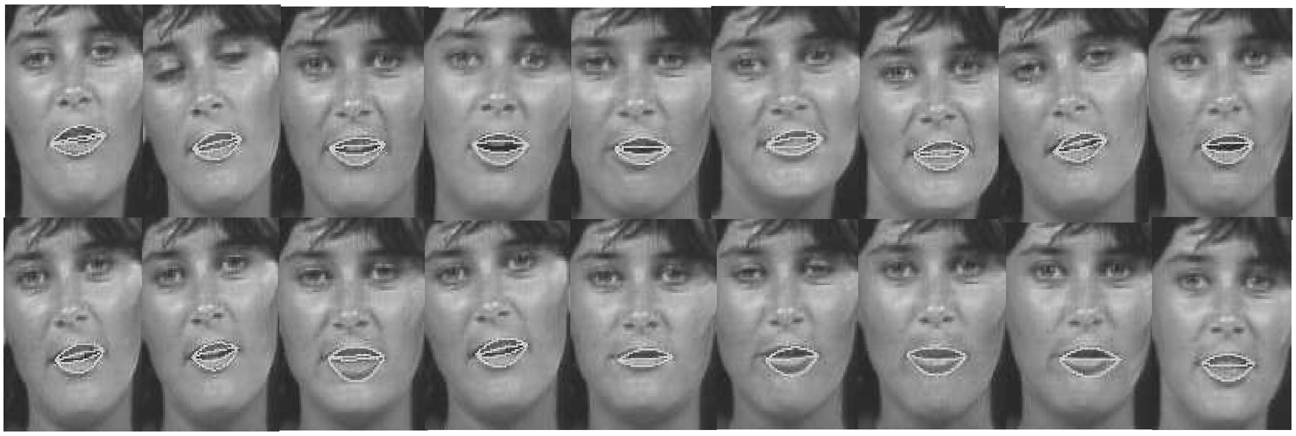


Fig. 3 Estimated mouth features in 18 frames for the sequence *Miss America* (CIF,10Hz).