Estimation of eye, eyebrow and nose features in videophone sequences

Markus Kampmann, Liang Zhang
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, 30167 Hannover, F.R.Germany
Email: {kampmann,zhang}@tnt.uni–hannover.de
WWW: http://www.tnt.uni–hannover.de/{~kampmann,~zhang}

Abstract

Automatic estimation of facial features is necessary for model–based coding of videophone sequences at very low bit rates. In this contribution, algorithms for the estimation of eye, eyebrow and nose features are presented. For estimation of eye features, deformable template matching with a simplified cost function is used. For the estimation of eyebrow features, a segmentation algorithm is proposed. Eyebrows covered by hair are taken into account by this algorithm. As nose features, the nostrils as well as the sides of the nose are estimated. Applied to the test sequences Akiyo *and* Miss America *(CIF,10 Hz), the facial features are estimated with subjectively high accuracy.*

1. Introduction

The automatic estimation of facial features like eyes, mouth, eyebrows, nose and face contours is necessary for many applications [1][2]. For example, facial feature estimation is applied to face recognition [1] and model–based coding of videophone sequences at very low bit rates [3][4][5]. This contribution deals with the estimation of eye, eyebrow and nose features in typical head–and–shoulder videophone sequences for very low bit rate video coding.

In the literature, several algorithms have been proposed for the estimation of eye, eyebrow and nose features. Deformable template matching as proposed in [6] has been frequently applied for the estimation of eye features [7][8]. In these algorithms, eye features are represented by a deformable eye template and this eye template is matched to a person's eye by minimizing a cost function. However, there are two problems associated with these approaches: (1) The cost function includes constraints about the interdependance between the different eye features, e.g. eye width should be six times as long as eye opening's height. These constraints are rarely fulfilled in typical videophone sequences; (2) The minimization of the cost function is performed in several stages and in each stage weighting factors within the cost function need different values. For example, more than 20 values are needed for 10 weighting factors in [6]. Eyebrows are often segmented assuming that the eyebrows are darker than the surrounding skin [9][10]. Other approaches use active contour models (*snakes*) [11] or assume high values of the luminance gradient at the eyebrow borders [12]. In many cases, eyebrows are partially or completely covered by the person's hair. This is not considered in the above mentioned algorithms. As nose features, the nose region is estimated in [13] and the nostrils are estimated in [9][10][11][14][15]. In [12], the sides of the nose are estimated in frontal face images for face recognition purposes.

In this contribution, new algorithms for the estimation of eye, eyebrow and nose features are proposed. These algorithms assume that eye and mouth center positions are already estimated with the algorithms in [16][17]. Eye features are represented by a deformable eye template. As eye features pupil (center of iris), eye corner points and eye opening's height are estimated. Here, a cost function without constraints about the eye features is introduced. For the estimation of eyebrow features, a segmentation algorithm is proposed. Eyebrows covered by hair are also taken into account by this algorithm. For the estimation of nose features, an algorithm for estimating the nostrils as well as the sides of the nose is proposed.

2. Estimation of eye features

Eye features are represented by a deformable eye template which is shown in Fig. 1. The deformable eye template consists of two parabels p_1 und p_2 and a circle c. The parabels represent the eyelids and the circle represents the iris. The parameters of the parabels and the circle describe the pupil (center of the iris), the radius of the iris, the eye corner points and the eyelid opening's height. The opening's height is determined using the maximal distance between the eyelids. Here, the pupil, the eye corner points and the eyelid opening's height are estimated. The radius of the iris is not estimated. It is approximated in proportion to the distance between the estimated eye corner points. The proposed algorithm for the estimation of eye features consists of three steps. In the first step, the pupil and the corner points of the eyes are estimated as described in [17][18]. These estimates are then used to define appropriate search windows for the estimation of the eyelid opening's height. In the second step, the eyelid opening's height is estimated using deformable template matching. For this, a simplified cost function is introduced, in which no assumption about the interdependance between the parameters is introduced. The cost function consists of only 2 terms considering spatial luminance gradients along eyelids and iris edges as well as means and variances of the luminance in the areas of eye white and iris. The minimization of this cost function is carried out in two steps. At first, candidates for eyelid opening's height are determined by only considering the luminance gradients. The eyelid opening's height is then estimated from these candidates by minimizing the cost function. During the minimization of this cost function the weighting factors are kept constant. In the final step, criteria are exploited to verify the estimated eyelid opening's height. If the estimated values do not pass the verification, they are rejected and the processing will be restarted in the subsequent image.

3. Estimation of eyebrow features

Due to the known eye centers [16][17], search areas for the eyebrows above the eye centers are introduced. Inside these search areas, the eyebrows are estimated assuming that the eyebrows are darker than the surrounding skin. Three different cases are taken into account: (1) The eyebrow is almost completely covered by hair; (2) A minor part of the eyebrow is covered by hair; (3) The eyebrow is not covered by hair. The estimation of the eyebrow features is carried out in three steps. First, it is detected whether a major part of the eyebrow is covered by hair (case 1). If a major part is covered by hair, no eyebrow features are estimated and the algorithm stops. Otherwise, a segmentation is carried out in the second step by binarizing the luminance image inside the search area. After binarization, it is automatically distinguished between case 2 and case 3. If the eyebrow is not covered by hair (case 3), the estimation is finished. If a minor part of the eyebrow is covered by hair (case 2), hair will be included in the segmented area after the second step. In this case, the third step is carried out. Here, it is detected first which part of the segmented area is the uncovered part of the eyebrow and which part is hair. Then, the part of the eyebrow covered by hair is recovered according to a model of a typical eyebrow.

4. Estimation of nose features

As nose features, the nostrils and the sides of the nose are estimated in two steps. In the first step, the nostrils are estimated. Since the mouth center is already known [16][17], search areas for the nostrils are introduced above the mouth. For the estimation of the nostrils it is assumed that at the border of a nostril the luminance gradient has a high value. Using this assumption, potential nostrils are detected in each search area. If a person is looking down, the nostrils will be not visible. Therefore, a verification of the potential nostrils is necessary. Only if the gradient values at the position of the potential nostrils are larger than a given threshold, the potential nostril is accepted. Otherwise, it is assumed that the person is looking down and the potential nostril is rejected. In the second step, the sides of the nose are estimated. Search areas for the sides are determined using the known mouth center [16][17]. If the nostrils are estimated in the first step, these search areas will be further reduced. Since the edge of the sides of the nose has a specific shape, a template of the side of the nose is introduced. Then, the correlation between this template and a luminance gradient image is computed. The points with the highest correlation are chosen as the sides of the nose.

5. Experimental results

The described algorithms have been applied to the test sequences *Akiyo* and *Miss America* (CIF,10 Hz). Fig. 2 and Fig. 3 show representative results for the estimation of eye, eyebrow and nose features, respectively. For the test sequence *Akiyo*, the nostrils are not estimated because the woman is looking down and the nostrils are not visible. All other facial features are estimated with subjectively high accuracy (Fig. 2). For the test sequence *Miss America*, the left eyebrows of the woman are not estimated since major parts are covered by hair. The right eyebrows are segmented because only minor parts are covered by hair. Also for *Miss America*, the facial features are estimated with subjectively high accuracy (Fig. 3).

6. Conclusions

Algorithms for the estimation of eye, eyebrow and nose features in head–and–shoulder videophone sequences have been introduced in this contribution. For the estimation of eye opening's height, deformable template matching with a simplified cost function is used. The cost function includes only two weighting factors which remain constant during minimization. For the estimation of eyebrow features, a segmentation algorithm is proposed. This algorithm takes into account eyebrows covered by hair. For the estimation of nose features, an algorithm is proposed estimating the nostrils as well as the sides of the nose. Applied to the test sequences *Akiyo* and *Miss America* (CIF,10 Hz), the facial features are estimated with subjectively high accuracy. Combined with algorithms for the estimation of mouth features [19] and face contours [20], it is now possible to estimate all necessary facial features for model–based video coding.

7. Acknowledgement

The authors wish to thank Dipl.–Ing. Vorapong Chanawatr for software support and helpful discussions on this work.

8. References

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

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] K.–M. Lam, H. Yan, "Locating and extracting the eye in human face images", *Pattern Recognition*, Vol. 29, No. 5, May 1996, pp. 771–779.

[8] X. Xie, R. Sudhakar, H. Zhuang, "On improving eye feature extraction using deformable templates", *Pattern Recognition*, Vol. 27, No. 6, June 1994, pp. 791–799.

[9] F. Lavagetto, S. Curinga, "Object–oriented scene modeling for interpersonal video communication at very low bit–rate", *Signal Processing: Image Communication*, Vol. 6, No. 5, October 1994, pp. 379–398.

[10] 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.

[11] C. Huang, C. Chen,"Human facial feature extraction for face interpretation and recognition", *Pattern Reconition*, Vol. 25, No. 12, 1992, pp. 1435–1444.

[12] R. Brunelli, T. Poggio, "Face recognition: features versus templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, October 1993, pp. 1042–1052.

[13] P. Antoszczyszyn, J. Hannah, P. Grant, "Facial features motion analysis for wire–frame tracking in model–based moving image coding", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 97)*, Munich, Germany, April 1997, pp. 2669–2672.

[14] E. Petajan, J. Zhang,"Robust talking face analysis for face animation and speech recognition", *International Workshop on Synthetic–Natural Hybrid Coding and Three Dimensional Imaging (IWSNHC3DI'97)*, Rhodes, Greece, September 1997, pp. 220–223.

[15] R. Stiefelhagen, J. Yang, "Gaze tracking for multimodal human–computer interaction", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 97)*, Munich, Germany, April 1997, pp. 2617–2620.

[16] 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.

[17] L. Zhang, "Tracking a face for knowledge–based coding of videophone sequences", *Signal Processing: Image Communication*, Vol. 10, No. 1–3, July 1997, pp. 93–114.

[18] 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.

[19] L. Zhang, "Estimation of the mouth features using deformable templates", *International Conference on Image Processing (ICIP–97)*, Santa Barbara, CA, October 1997, pp. 328–331.

[20] M. Kampmann, "Estimation of the chin and cheek contours for precise face model adaptation ", *International Conference on Image Processing (ICIP–97)*, Santa Barbara, CA, October 1997, pp. 300–303.

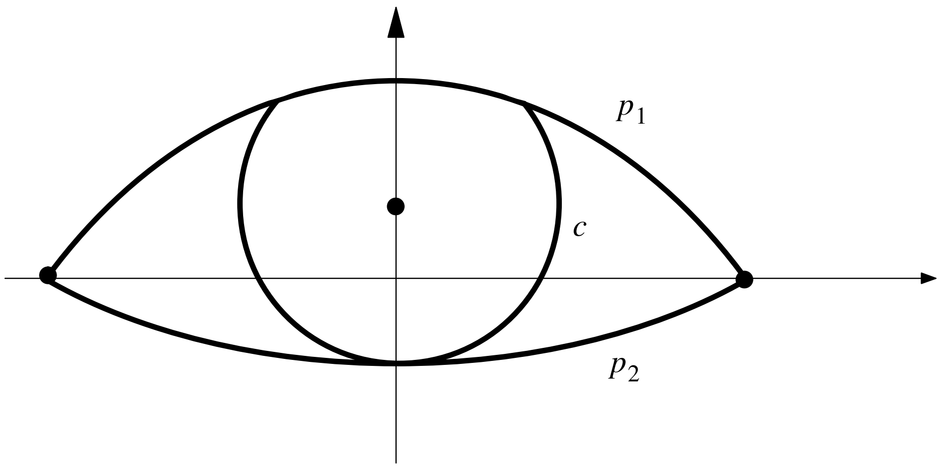

Fig. 1: Deformable eye template.

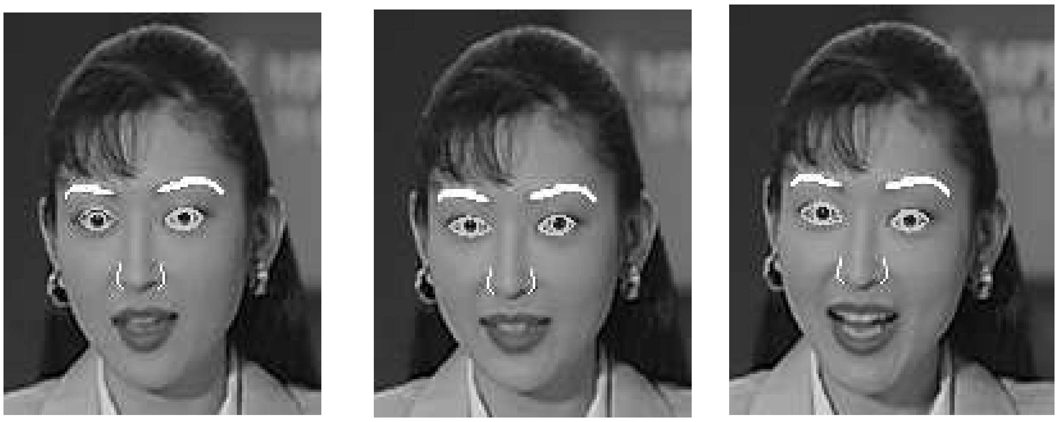

Fig. 2: Test sequence *Akiyo* (CIF, 10Hz): estimated facial features.

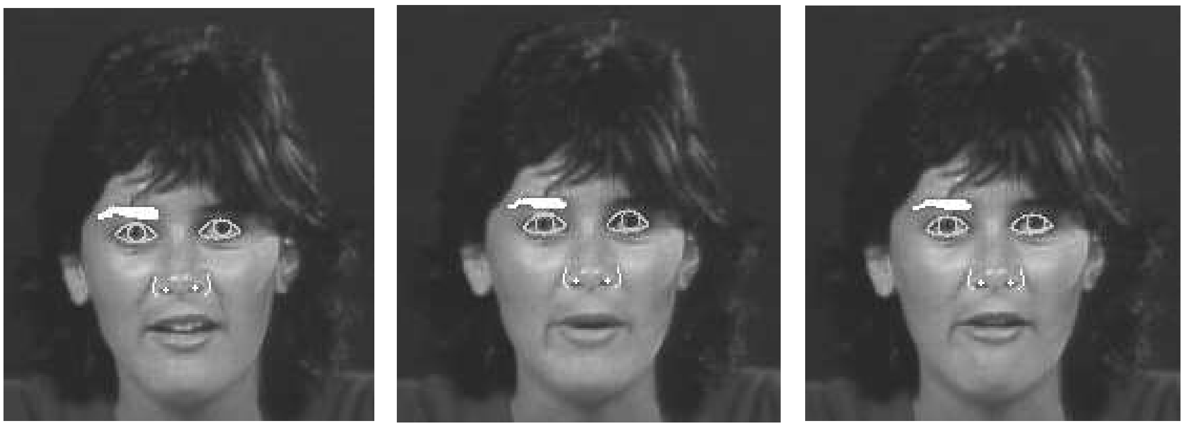

Fig. 3: Test sequence *Miss America* (CIF, 10Hz): estimated facial features.