Estimation of eye and mouth corner point positions in a knowledge–based coding system

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

Abstract

Automatic extraction of facial feature points is one of the main problems for semantic coding of videophone sequences at very low bit rates. In this contribution, an approach for estimation of the eye and mouth corner point positions is presented. For this proposition, the location informations of the face model are exploited to define search areas for estimation of the eye and mouth corner point positions. Then, the eye and mouth corner point positions are estimated based on a template matching technique with eye and mouth corner templates. Finally, in order to verify these estimated corner point positions, some geometric conditions between the corner point positions and the center point positions of the eyes and the mouth are exploited. The proposed algorithm has been applied to test sequences *Claire* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz.

Keywords: Facial feature points, eye and mouth corner point positions, face model, template matching.

1. Introduction

For coding of head–and–shoulder videophone sequences at very low bit rates, the knowledge of the presence of a human face can be exploited by a knowledge–based coder[1]. If the knowledge about the facial expressions in the secene is available, it can be exploited by extending a knowledge–based coder to a semantic coder[2], too. For implementation of a semantic coding system[3], a face model must be automatically adapted not only to a person's face in the sequence but also to its facial expressions. Some algorithms for automatically adapting the face model "CANDIDE" to the person's face based on the center points of the eyes and the mouth has been proposed[4,1]. For the further adaptation of the face model to facial expressions, 28 structural feature points on a person's face in the image have to be automatically estimated first[3]. In this contribution, only estimation of the eye and mouth corner point positions is addressed.

Many approaches for automatic extraction of the facial feature points have been proposed. An overview about these approaches can be found in the papers[5,6]. Some methods for corner detection in a curve have been developed for computer vision applications[7,8]. In the paper[8], corner detection is cast as a problem of cost optimization and the cost function captures different desirable characteristics of corners such as

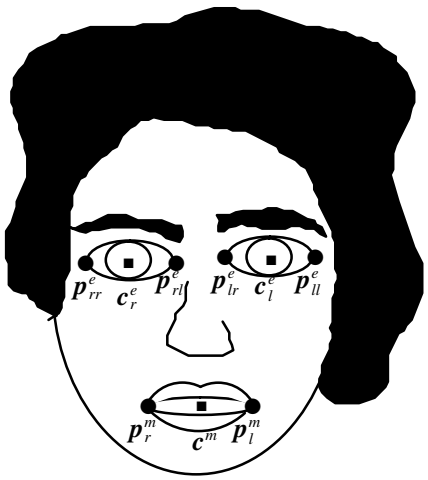
edginess, curvature and region dissimilarity. In the paper[9], this corner detection approach[8] is used to detect the eye corner points in order to reduce the processing time for the deformable template[10]. But, the performance of this corner detection depends on the gray–level contrast around the eye.

In this contribution, an approach for estimation of the eye and mouth corner point positions is presented. After the face model is tracked frame by frame in a knowledge–based coding system[11], the center point positions of the eyes and the mouth of the face model projected onto the image plane already match those positions of the real person in the image. Therefore, the location informations of the face model are exploited to define search areas for estimation of the eye and mouth corner point positions. Then, the eye and mouth corner point positions are estimated based on a template matching technique with eye and mouth corner templates. Finally, in order to verify these estimated corner point positions, some geometric conditions between the corner point positions and the center point positions of the eyes and the mouth are exploited. The proposed algorithm has been applied to test sequences *Claire* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz.

This contribution is organized as follows. Section 2 discusses some geometric conditions of the facial feature points on a human face in the image plane. Section 3 describes estimation of the eye and mouth corner point positions in the image. The experimental results using real image sequences *Claire* and *Miss America* are given in Section 4.

2. Geometric conditions of the facial feature points

This algorithm is developed for the head–and–shoulder videophone sequences, which facial features, eyes and mouth, are visible on the image plane and the eyes are open. It is assumed, too, that the pupils of the eyes lie near the center positions of the eyes. Hence, the feature points, corner and center point positions of the eyes and the mouth (Fig. 1), have the following geometric conditions,



- Eye and mouth center point positions
  c^e_r c^e_l c^m
- Eye and mouth corner point positions
  p^e_{rr} p^e_{rl} p^e_{lr} p^e_{ll} p^m_r p^m_l

Figure 1. Facial feature points on a human face.

1. The corner point connecting lines $\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}$, $\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}$ and $\boldsymbol{p}^m_l\boldsymbol{p}^m_r$ are approximately parallel to the connecting line of both eye center points $\boldsymbol{c}^e_l\boldsymbol{c}^e_r$, respectively,
2. The length of the left eye $\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}$ is equal to that of the right eye $\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}$,
3. The length $\boldsymbol{c}^m\boldsymbol{p}^m_l$ is equal to that $\boldsymbol{c}^m\boldsymbol{p}^m_r$,

where $\boldsymbol{c}^e_l$, $\boldsymbol{c}^e_r$ and $\boldsymbol{c}^m$ stand for the center point positions and $\boldsymbol{p}^e_{lr}$, $\boldsymbol{p}^e_{ll}$, $\boldsymbol{p}^e_{rr}$, $\boldsymbol{p}^e_{rl}$, $\boldsymbol{p}^m_l$ and $\boldsymbol{p}^m_r$ stand for the corner point positions, respectively. These geometric conditions are exploited to verify the correctness of the estimated corner point positions of the eyes and the mouth.

3. Estimation of the eye and mouth corner point positions

The algorithm for estimation of the eye and mouth corner point positions consists of following steps:

Detection of potential areas for the eye and mouth corner point positions

Fig. 2 illustrates the diagram for detection of potential areas for the eye and mouth corner point positions. After the face model is tracked frame by frame[11], the center point positions of the eyes and the mouth of the face model projected onto the image plane already match those positions of the real person in the image. Therefore, this location informations of the face model are exploited to define search areas for estimation of the eye and mouth corner point positions. After that, in order to further reduce the search areas, the template matching is used. The corner templates used in this contribution were created from the test sequneces *Claire, Miss America* and *Michael*. The eye and mouth corners of the person in these sequences were extracted and changed to equal size. Then, the average eye and mouth corners were computed from these eye and mouth corners, respectively. Finally, these average eye and mouth corners are used as the eye and mouth corner templates. The templates are not only roughly adapted to the size of the person's face but also inclined in accordance with the inclination of the head. In each search area, the correlation between the real image and a corresponding corner template is computed in a window centered at (x,y). Points with high values of this correlation are extracted as potential area for the corner point position. The subsequent algorithms for estimation of the eye and mouth corner point positions are applied only to these potential areas which might contain the eye and mouth corner point positions.

Estimation of the eye and mouth corner point coordinates

For estimation of the corner point coordinates, a probability measure f is evaluated:

$$f = c^k_{factor}(x, y)\,, \tag{1}$$

where $c^k_{factor}(x, y)$ are the correlations between the real image k and an artificial corner template with a different size factor of 0.8, 0.9, ..., 1.4. The point with the highest values of f in each potential area is selected as 2D corner point coordinates.

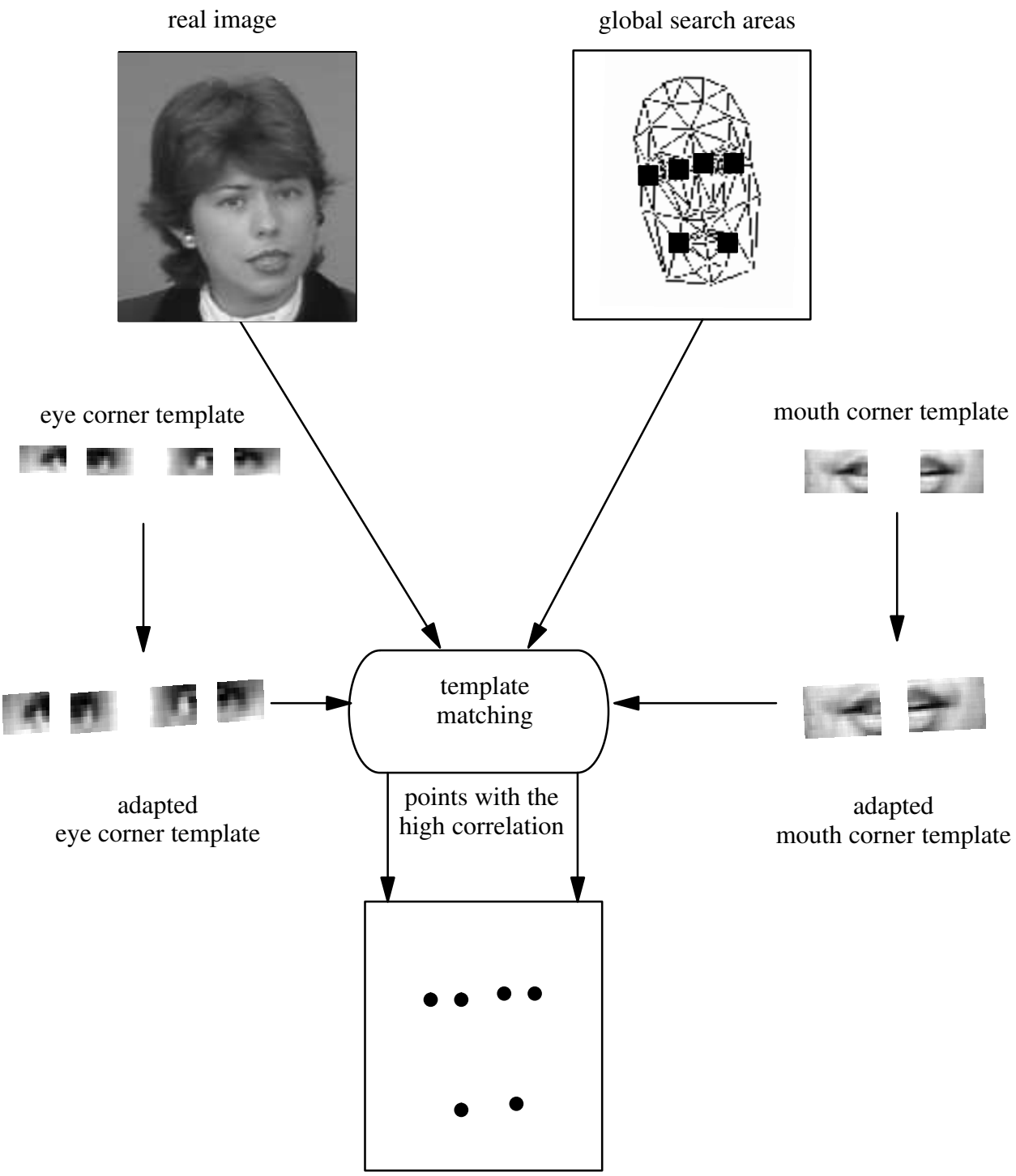


Figure 2. Detection of potential areas for the corner point positions.

Verification of the correctness of the estimated corner point coordinates

After the eye and mouth corner point coordinates have been estimated, their correctness has to be verified. According to the geometric conditions in Section 2, the following criteria are used for this goal.

1. The inclinations of the corner point connecting lines $\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}$, $\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}$, $\boldsymbol{p}^m_l\boldsymbol{p}^m_r$ and the inclination of the eye center point connecting line $\boldsymbol{c}^e_l\boldsymbol{c}^e_r$ must respectively satisfy

$$\left|\alpha_{\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}} - \alpha_{\boldsymbol{c}^e_l\boldsymbol{c}^e_r}\right| < th_1,$$

$$\left|\alpha_{\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}} - \alpha_{\boldsymbol{c}^e_l\boldsymbol{c}^e_r}\right| < th_1,$$

$$\left|\alpha_{\boldsymbol{p}^m_l\boldsymbol{p}^m_r} - \alpha_{\boldsymbol{c}^e_l\boldsymbol{c}^e_r}\right| < th_2,$$

   where inclination α is defined as

$$\alpha = \arctan\left[\frac{Y_l - Y_r}{X_l - X_r}\right].$$

2. Lengths of both eyes must satisfy

$$\frac{\left|L_{\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}} - L_{\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}}\right|}{\left|L_{\boldsymbol{p}^e_{rl}\boldsymbol{p}^e_{rr}} + L_{\boldsymbol{p}^e_{ll}\boldsymbol{p}^e_{lr}}\right|} < th_3.$$

3. A half of the mouth length must satisfy

$$\frac{\left|L_{\boldsymbol{p}^m_r\boldsymbol{c}^m} - L_{\boldsymbol{p}^m_l\boldsymbol{c}^m}\right|}{\left|L_{\boldsymbol{p}^m_r\boldsymbol{c}^m} + L_{\boldsymbol{p}^m_l\boldsymbol{c}^m}\right|} < th_4.$$

The thresholds th_1, th_2, th_3 and th_4 are considered constant during the sequence. If the above conditions are satisfied, these estimated coordinates are taken as the eye and mouth corner point coordinates, respectively.

4. Experimental results

The described algorithm was applied to the test sequences *Claire* and *Miss America* with a spatial resolution corresponding to CIF and a frame rate of 10 Hz. In order to measure the accuracy of the eye and mouth corner point estimates, the maximum position errors $f^{\max}_{eye\ corner,n}$, $f^{\max}_{mouth\ corner,n}$ over the frame n and their average position errors $\bar{f}_{eye\ corner}$, $\bar{f}_{mouth\ corner}$,

$$\bar{f}_{eye\ corner} = \frac{1}{N_{eye}} \sum_n f^{\max}_{eye\ corner,n}, \tag{2}$$

$$\bar{f}_{mouth\ corner} = \frac{1}{N_{mouth}} \sum_n f^{\max}_{mouth\ corner,n}, \tag{3}$$

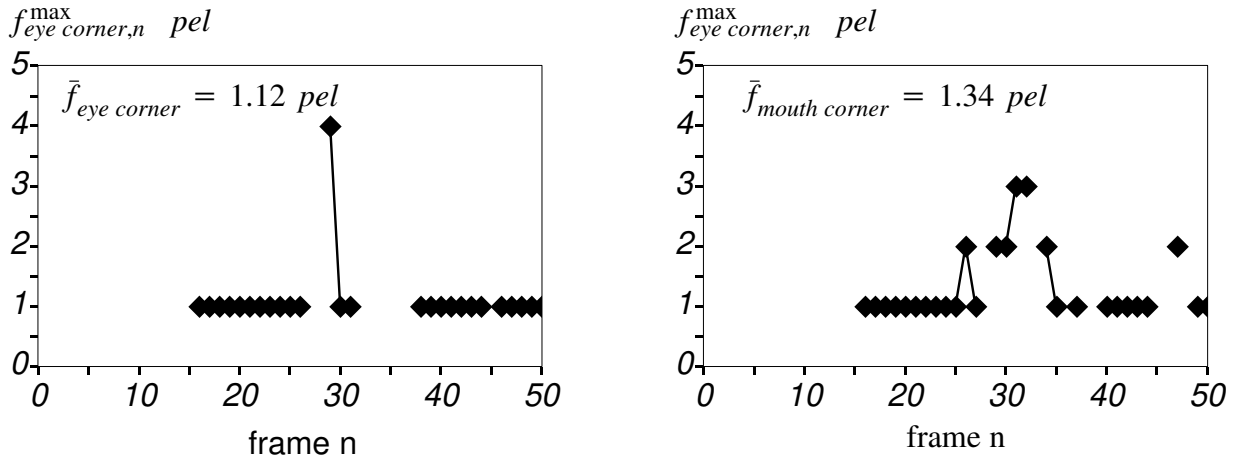
$$f^{\max}_{eye\ corner,n} = \max\left[|X'_{ll,n} - X^t_{ll,n}|, |Y'_{ll,n} - Y^t_{ll,n}|, |X'_{lr,n} - X^t_{lr,n}|, |Y'_{lr,n} - Y^t_{lr,n}|,\right.$$

$$|X'_{rl,n} - X^t_{rl,n}|, |Y'_{rl,n} - Y^t_{rl,n}|, |X'_{rr,n} - X^t_{rr,n}|, |Y'_{rr,n} - Y^t_{rr,n}|\Big], \quad (4)$$

$$f^{\max}_{mouth\ corner,n} = \max\Big[|X'_{ml,n} - X^t_{ml,n}|, |Y'_{ml,n} - Y^t_{ml,n}|, |X'_{mr,n} - X^t_{mr,n}|, |Y'_{mr,n} - Y^t_{mr,n}|\Big], \quad (5)$$

are introduced, where $(X^t_{ll,n} Y^t_{ll,n})$, $(X^t_{lr,n}, Y^t_{lr,n})$, $(X^t_{rl,n}, Y^t_{rl,n})$, $(X^t_{rr,n}, Y^t_{rr,n})$, $(X^t_{ml,n}, Y^t_{ml,n})$and $(X^t_{mr,n}, Y^t_{mr,n})$ are the true image corner coordinates of the eyes and the mouth at the image n (manually determined), $(X'_{ll,n}, Y'_{ll,n})$, $(X'_{lr,n}, Y'_{lr,n})$, $(X'_{rl,n}, Y'_{rl,n})$, $(X'_{rr,n}, Y'_{rr,n})$, $(X'_{ml,n}, Y'_{ml,n})$ and $(X'_{mr,n}, Y'_{mr,n})$ are the estimated corner coordinates of the eyes and the mouth at the image n. N_{eye} is the number of images in which the eye corner coordinates are estimated. N_{mouth} is the number of images in which the mouth corner coordinates are estimated. In this experiment, the threshold th_1, th_2, th_3 and th_4 are 15^0, 15^0, 0.15 and 0.2, respectively.

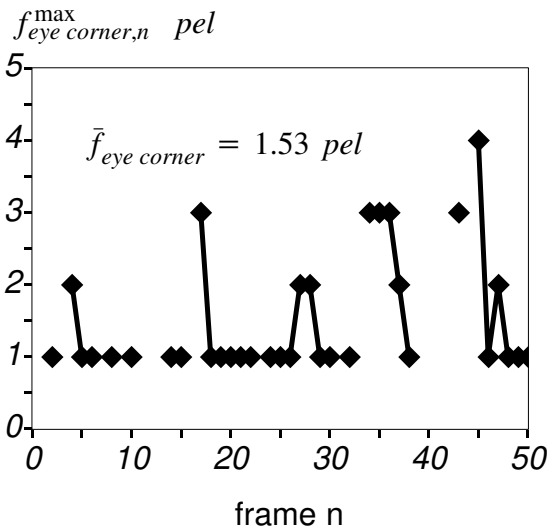
For the test sequences *Claire* (CIF, 10 Hz), the face model is adapted to the person's face in the 16th frame. 74% positions of the eye corner points and 77% positions of the mouth corner points are successfully estimated. Fig.3 shows the maximum position errors $f^{\max}_{eye\ corner,n}$ and $f^{\max}_{mouth\ corner,n}$ over the frame n. Their average position errors $\bar{f}_{eye\ corner}$ and $\bar{f}_{mouth\ corner}$ are 1.12 and 1.34 pel, respectively. Fig.4 shows the estimated eye and mouth corner point positions in the 26th frame.



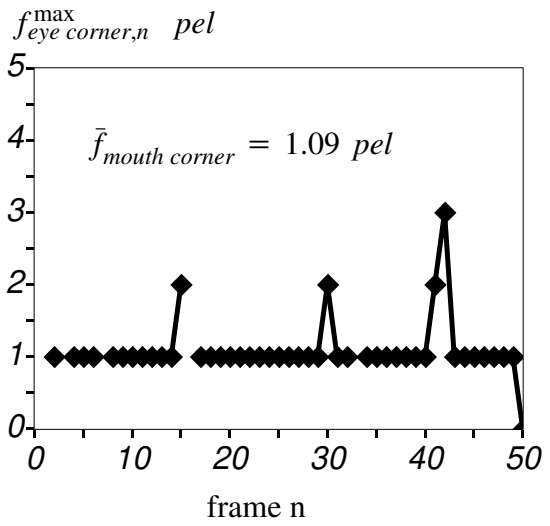
(a) Maximum position error for eye corner (b) Maximum position error for mouth corner

Figure 3. Maximum position errors for the eye and mouth corner positions for test sequence *Claire* (CIF, 10Hz).

For the test sequences *Miss America* (CIF, 10 Hz), the face model is adapted to the person's face in the second frame. 69% positions of the eye corner points and 91% positions of the mouth corner points are successfully estimated. Fig.5 shows the maximum position errors $f^{\max}_{eye\ corner,n}$ and $f^{\max}_{mouth\ corner,n}$ over the frame n. Their average position errors $\bar{f}_{eye\ corner}$ and $\bar{f}_{mouth\ corner}$ are 1.53 and 1.09, respectively. Fig.6 shows the estimated eye and mouth corner point positions in the 27th frame.



(a) Maximum position error for eye corner



(b) Maximum position error for mouth corner

Figure 5. Maximum position errors for the eye and mouth corner positions for test sequence *Miss America* (CIF, 10Hz).

Figure 4. The estimated eye and mouth corner positions in the 26th frame of the test sequence *Claire* (CIF, 10Hz).

Figure 6. The estimated eye and mouth corner positions in the 27th frame of the test sequence *Miss America* (CIF, 10Hz).

5. Conclusion

In this contribution, an algorithm for estimation of the eye and mouth corner point coordinates has been proposed. After the face model is tracked frame by frame in a knowledge–based coding system, the eye and mouth corner point coordinates of a person in the image are estimated. For both test sequences *Claire* and *Miss America,* the average position errors for the eye corner and the mouth corner are found to be 1.33 and 1.22 pel, respectively.

References

[1] 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", accepted for publication in *Signal Processing: Image Communication*, Special Issue on Communicating Natural Scenes by Image Synthesis.

[2] 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.

[3] 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.

[4] R. Koch, "Adaptation of a 3D facial mask to human faces in videophone sequences using model based image analysis", *Picture Coding Symposium* (PCS'91), Tokyo, Japan, pp. 285–288, September 1991.

[5] A. Samal and 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.

[6] R. Chellappa, C.L. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", *Proceedings of IEEE*, Vol. 83, No. 5, May 1995, pp. 705–740.

[7] C. Teh and R. Chin, "On the detection of dominant points on digital curves", *IEEE Trans. Pattern Anal. Mach. Comput.* Vol. 11, No. 8, August 1989, pp. 859–872.

[8] X. Xie, R. Sudhakar and H. Zhuang, "Corner Detection by a cost minimization approach", *Pattern Recognition*, Vol. 26, No. 8, August 1993, pp. 1235–1243.

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

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

[11] L. Zhang, "Tracking a face in a knowledge–based analysis–synthesis coder", *International Workshop on Coding Techniques for Very Low Bit–Rate Video* (VLBV 95), Tokyo, Japan, paper A–6, November 1995.