FACIAL FEATURE EXTRACTION BASED ON THE SMALLEST UNIVALUE SEGMENT ASSIMILATING NUCLEUS (SUSAN) ALGORITHM

Mauricio Hess[1] *Geovanni Martinez*[2]

Image Processing and Computer Vision Research Lab (IPCV-LAB)
Escuela de Ingenieria Electrica, Universidad de Costa Rica
Codigo postal 2060, San Jose, Costa Rica
[1]mauhf@costarricense.cr, [2]gmartin@pacuare.eie.ucr.ac.cr

ABSTRACT

For coding of videophone sequences at very low bit rates, model-based coding is investigated. In a model-based coder, the human face in the videophone sequence is described by a three-dimensional (3D) face model. At the beginning of the videophone sequence, the model has to be adapted automatically to the size, position and orientation of the real face present in the scene. First, the face of the person is segmented in the current image. Then, facial features such as eye corners and centers, mouth corners and center, chin and cheek borders, nose corners, etc. are extracted. In the final step the model is adapted using the extracted facial features. In this paper, a new facial feature extraction algorithm is presented. First, all corners and borders are extracted using the Smallest Univalue Segment Assimilating Nucleus (SUSAN) Algorithm. Then, the facial features are detected by applying knowledge-based rules. The algorithm was implemented and applied to the first 40 frames of the videophone sequence *Claire* (CIF, 10Hz) with very encouraging results.

1. INTRODUCTION

For coding of videophone sequences at very low bit rates, model-based coding is investigated [1, 2, 3, 4, 5, 6]. In a model-based coder, the human face in the videophone sequence is described by a three-dimensional (3D) face model. The shape is represented by a 3D wire-frame whose vertices represent the shape parameters. The motion is described by six parameters: three rotation angles and one 3D translation vector [7, 8]. The mimic is described by action units [9] or facial muscle parameters [10]. The color is described by projecting a real image onto the wire-frame. These parameters are estimated at the sender [10, 11, 12, 13, 14], then coded and transmitted to the receiver. After decoding of the transmitted parameters at the receiver, a synthetic image of the face is generated.

Usually, a generic 3D face model is used. This model has the proportions of an average human face. At the beginning of the videophone sequence, the model has to be adapted automatically to the size, position and orientation of the real face present in the scene [15]. First, the face of the person is segmented in the current image. Then, facial features such as the eye corners and centers, mouth corners and center, chin and cheek borders, nose corners, etc. are extracted. In the final step the model is adapted using the extracted facial features.

Some algorithms for human face segmentation have been proposed in the professional literature, such as those described in [16, 17], where the face is found using color segmentation. In [18] and [19], ellipses are adjusted to a border image to segment the face. In [20], the previously estimated silhouette of the person is subdivided into head and shoulders and then the upper third of the head area is removed to segment the face.

In the professional literature, some algorithms are proposed for the extraction of facial features such as the eye corners and centers, the mouth corners and center, and the nose corners. Many of these are based on a method known as artificial template matching [20, 21, 22, 23, 24, 25, 26]. An artificial template is a small rectangular intensity image that contains for example a mouth corner, where the corner is located in the center of the artificial template. The image region which best matches the artificial template is extracted from the current image. To generate an artificial template, for example, of a mouth corner, rectangular regions that contain the corner are first manually extracted from intensity images of different people. Then the regions are scaled so that they all have the same size and finally their intensities are averaged to obtain the artificial template. The problem with artificial template matching is that it fails when it is applied to images that are different from those that were used to generate the artificial templates.

For the extraction of the chin and cheek borders, a method known as deformable template matching is used [26, 27, 28, 29, 30]. First, a region in the current image which probably contains the border to be extracted is segmented using the previously extracted positions of the eye and mouth corners

and centers. Finally, assuming that the pixels on the real border have high linear gradients, the border is extracted by deforming one or more parabolas (deformable templates) until they match most of the pixels with high linear gradients.

In this paper, a new facial feature extraction algorithm is proposed. The most significant contribution is that the artificial template matching and the deformable template matching will not be used to extract facial features. For the extraction of the outer corners of the eyes, the mouth corners and the nose corners, the SUSAN (Smallest Univalue Segment Assimilating Nucleus) algorithm for corner and border extraction [31] will be used as an alternative to artificial template matching. For the extraction of the chin and cheek borders, the SUSAN algorithm for border extraction [31] will be used instead of deformable template matching.

The proposed algorithms will be applied to real videophone sequences to compare their reliability and accuracy with existing algorithms in the professional literature.

This paper is organized as follows. In section 2, the extraction algorithm of the mouth corners and outer corners of the eyes is described. In section 3, the extraction algorithm of the chin and cheek borders is presented. In section 4, the extraction algorithm of the nose corners is described. In section 5 and in section 6, experimental results and the conclusions are given, respectively.

2. EXTRACTION OF THE MOUTH CORNERS AND OUTER CORNERS OF THE EYES

The first step consists of the establishment of a rectangular search region for the mouth and a rectangular search region for the eyes. First, the face in the current image is segmented by applying the method described in [20], where the previously estimated silhouette of the person is subdivided into head and shoulders assuming that the shoulders are wider than the head. The middle third of the segmented head represents a non-rectangular region that contains the eyes (see Fig. 1(a)), and the lower third of the segmented head represents a non-rectangular region that contains the mouth (see Fig. 1(a)). The rectangular search region of the mouth is defined as the smallest rectangular region that encloses the previously found region that contains the mouth (see Fig. 1(b)), and the rectangular search region of the eyes is defined as the smallest rectangular region that encloses the previously found region that contains the eyes (see Fig. 1(b)).

In the next step, the size of the rectangular search region of the mouth is reduced and then this re-sized region is subdivided vertically along its middle column m into two rectangular halves corresponding to the rectangular search region of the left mouth corner (see Fig. 2(a)) and the rectangular search region of the right mouth corner (see Fig. 2(b)).

(a) (b)

Fig. 1. (a) Middle third and lower third of the segmented head. (b) Rectangular search regions for the mouth and both eyes.

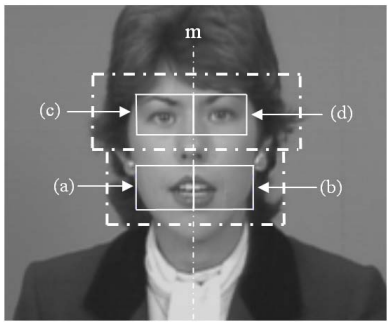


Fig. 2. Rectangular search regions for the mouth corners and outer corners of the eyes.

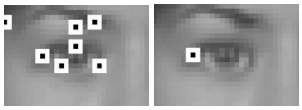
The re-sized search region of the mouth is that rectangular region within the original one that has the highest border concentration. The borders are extracted by applying the SUSAN border extraction algorithm [31] to the intensity component of the current image.

In the following step, the size of the rectangular search region of the eyes is reduced and then this re-sized region is subdivided vertically along the column m into the rectangular search region of the outer corner of the left eye (see Fig. 2(c)) and the rectangular search region of the outer corner of the right eye (see Fig. 2(d)). The re-sized search region of the eyes is that rectangular region within the original one that has the highest border concentration.

In the next step, the positions of the mouth corners are extracted. Since the extraction algorithm of the right mouth corner is similar to the extraction algorithm of the left mouth corner, only the latter is described. First, the SUSAN algorithm [31] for corner extraction is applied to the intensity component of the current image, inside the rectangular search region for the left mouth corner (see Fig. 3(a)). Usually, more than one corner is extracted. The extracted corners are called here left mouth corner candidates. To detect the correct left mouth corner from the candidates three rules are applied one after the other. First, a candidate is eliminated, if the amount of columns between it and the column m is greater than 35% of the width of the search region. Second, a remaining candidate is eliminated if the border

(a) (b) (c)

Fig. 3. (a) Extracted candidate corners for the left mouth corner. (b) Extracted borders. (c) Detected left mouth corner.



(a) (b)

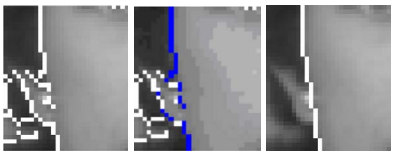
Fig. 4. (a) Extracted candidate corners for the outer corner of the left eye. (b) Detected outer corner of the left eye.

concentration to its right is lower than the border concentration to its left. Borders are extracted by applying the SUSAN algorithm [31] to the intensity component of the current image, inside the rectangular search region for the left mouth corner (see Fig. 3(b)). Third, that remaining candidate which is located most to the left of column m is considered to be the left mouth corner (see Fig. 3(c)).

In the following step, the positions of the outer corners of the eyes are extracted. Since the extraction algorithm of the outer corner of the right eye is similar to the extraction algorithm of the outer corner of the left eye, only the latter is described. First, the SUSAN algorithm for corner extraction is applied to the intensity component of the current image, inside the rectangular search region for the outer corner of the left eye. The extracted corners are called here candidates for the outer corner of the left eye (see Fig. 4(a)). To detect the correct outer corner of the left eye from the candidates three rules are applied one after the other. First, a candidate is eliminated if the border concentration to its right is lower than the border concentration to its left. Borders are extracted by applying the SUSAN algorithm for border extraction to the intensity component of the current image, inside the rectangular search region for the outer corner of the left eye. Second, beginning from the column m to its left, the distance between a candidate and each one of the candidates to its right is measured, the candidate is eliminated if all the measured distances are greater than $1/5$ of the width of the search region. Third, the distance between the middle of the lower eyelid and each remaining candidate above and to its left is measured, the outer corner of the left eye is that candidate which is most to the left and whose distance is no greater than $1/3$ of the width of the mouth (see Fig. 4(b)). The middle of the lower eyelid is that remaining candidate whose position is the lowest inside the search region.

(a) (b)

Fig. 5. (a) Segmented left cheek border region and segmented right cheek border region. (b) Rectangular search regions for the chin and cheek borders.



(a) (b) (c)

Fig. 6. (a) Extracted borders. (b) Selected borders which are most to the right. (c) Extracted left cheek border.

3. EXTRACTION OF THE CHIN AND CHEEK BORDERS

The first step in the extraction of the chin and cheek borders consists of the establishment of rectangular search regions for the left cheek border, the right cheek border, and the chin border. Since the algorithm for the establishment of the rectangular search region of the right cheek border is similar to the algorithm for the establishment of the rectangular search region of the left cheek border, only the latter is described. First, a region that contains the left cheek border is segmented. The segmented left cheek border region is that part of the silhouette which is to the left of the column which intersects the extracted left mouth corner and between the lowest row of the segmented head and the row that intersects the extracted outer corner of the left eye (see Fig. 5(a)). The rectangular search region of the left cheek border is defined as the smallest rectangular region that encloses the segmented left cheek border region (see Fig. 5(b)). The rectangular search region of the chin border is defined as that rectangular region enclosed by the column that intersects the extracted right mouth corner, the column that intersects the extracted left mouth corner, the lowest row of the segmented head and that row c which is $1/6$ of the height of the segmented head below the lowest row of the segmented head (see Fig. 5(b)).

In the next step, the chin and cheek borders are extracted from the rectangular search regions previously found. Since the extraction algorithms of the right cheek border and the chin border are similar to the extraction algorithm of the left cheek border, only the latter is described. First, the

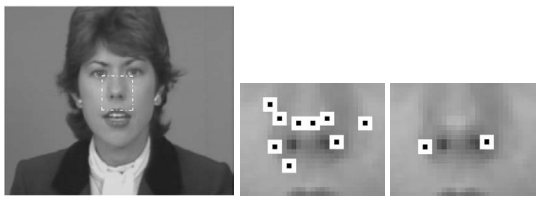
SUSAN algorithm for border extraction is applied to the intensity component of the current image, inside the rectangular search region of the left cheek border, as shown in Fig. 6(a)). Unfortunately, other borders are extracted, corresponding especially to hair near the cheeks. To detect the correct left cheek border from the extracted borders three rules are applied one after the other. First, assuming that no borders are extracted inside the left cheek, those borders which are most to the right inside the search region are selected, (marked in Fig. 6(b)). This usually results in non-soft border sections, i.e. border sections with drastic direction changes. Second, pixels of each border section where the direction changes drastically are eliminated. Third, the resulting soft border sections are connected by adding pixels in such a way that one continuous border with no drastic direction changes is obtained. This resulting border represents the left cheek border (see Fig. 6(c)).

4. EXTRACTION OF THE NOSE CORNERS

The first step in the extraction of the right nose corner and the left nose corner consists of the establishment of a rectangular search region that contains both corners. This region is defined as that rectangular region enclosed by the column that intersects the extracted left mouth corner, the column that intersects the extracted right mouth corner, the row that intersects the highest of the extracted mouth corners and the row that intersects the lowest of the extracted outer corners of the eyes (see Fig. 7(a)). In the next step, the nose corners are extracted. Since the extraction algorithm of the right nose corner is similar to the extraction algorithm of the left nose corner, only the latter is described. First, the SUSAN algorithm for corner extraction is applied to the intensity component of the current image, inside the rectangular search region that contains both nose corners (see Fig. 7(b)). Unfortunately, besides the nose corners others are extracted. All the extracted corners are called here nose corner candidates. To detect the correct left nose corner from the candidates, two rules are applied one after the other. First, the distance between each candidate and all the other candidates is measured; the candidate is eliminated if all the distances are greater than $1/4$ of the width of the search region. Second, the distance between the remaining candidate which is most to the left and each remaining candidate below it is measured, the left nose corner is that candidate which is most to the right and whose distance is no greater than $1/6$ of the width of the search region. The detected nose corners are shown in Fig. 7(c).

5. EXPERIMENTAL RESULTS

The proposed algorithm was implemented in C and applied to the first 40 frames of the test sequence *Claire* (CIF, 10 Hz). Using an Intel Pentium IV processor at 1.6 GHz, with 256 MB of RAM and under Windows XP, the mean processing time was 0.197 s. For SUSAN corner extraction [31] the used intensity level threshold 't' was 22 for the mouth corners and the outer corners of the eyes, and 9 for the nose corners. For SUSAN border extraction [31] the used intensity level 't' was 15 for the mouth corners, 21 for the outer corners of the eyes and the cheek borders, and 12 for the chin border.



(a) (b) (c)

Fig. 7. (a) Rectangular search region that contains both nose corners. (b) Candidate corners for the nose corners. (c) Detected nose corners.

The reliability and accuracy of the proposed algorithm were measured to evaluate its performance. To measure the reliability, the percentage of images where the algorithm fails was obtained. It is defined that the algorithm fails when at least one of the facial features (mouth corners, outer corners of the eyes, nose corners or chin and cheek borders) cannot be extracted. According to the experimental results, the algorithm did not fail in 85% of the analyzed frames. To measure the accuracy of the proposed algorithm, the extracted facial features were compared with manually extracted facial features. The obtained average position errors were 2.252 ± 0.861 pixels for the mouth corners and outer corners of the eyes, 3.655 ± 1.854 pixels for the nose corners, and 1.097 ± 0.744 pixels for the chin and cheek borders. Some examples of the obtained results are depicted in Fig. 8.

The reliability and precision of the proposed feature extraction algorithm were compared to those of other algorithms from the professional literature. In [25], artificial template matching is applied to extract the eye and mouth corners in frames of the test sequences *Claire* and *Miss America* using artificial templates created from the test sequences *Claire*, *Miss America* and *Michael*. For *Claire*, the reliability percentages of the eye corner extraction and the mouth corner extraction were 74% and 77%, respectively. The average of the maximum position errors were 1.12 pixels for the eye corners and 1.34 pixels for the mouth corners. The reliability and accuracy of the mentioned algorithm would probably be lower if it is applied to frames of sequences other than those used to create the artificial templates. Using the proposed algorithm, the positions of the mouth corners and outer corners of the eyes were extracted with a reliabil-

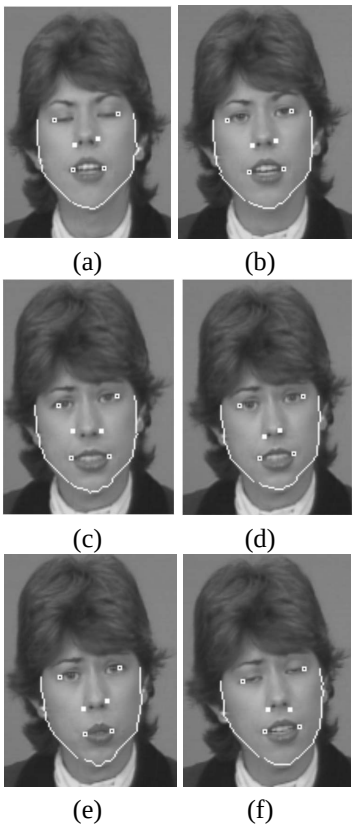


Fig. 8. Examples of the extracted facial features for frames of the test sequence *Claire*. (a) 4^{th} frame, (b) 7^{th} frame, (c) 12^{th} frame, (d) 14^{th} frame, (e) 17^{th} frame, (f) 29^{th} frame.

ity percentage of 90% and a average position error of 2.25 pixels.

In [29], deformable template matching is used to extract the chin and cheek borders, for the first 50 frames of the test sequences *Akiyo* and *Claire*. The chin and cheek borders were extracted with no position error for more than 60% of the analyzed frames. In the remaining frames, only small position errors were noticed. Using the algorithm for the extraction of the chin and cheek borders proposed in this paper, the chin and cheek borders were extracted with a reliability percentage of 85% and a mean position error of 1.09 pixels.

6. SUMMARY AND CONCLUSIONS

A new algorithm for facial feature extraction in videophone sequences has been presented which is based on the SUSAN algorithm for corner and border extraction. Since the SUSAN algorithm also extracts other here irrelevant facial features (hair, teeth, eyebrows, etc.) some knowledge-based rules are applied to detect among them the mouth corners, outer corners of the eyes, nose corners, as well as the chin and cheek borders. The main contribution of this paper is that the new algorithm is based on corner and border extraction instead of being based on artificial template matching and deformable template matching. The algorithm was applied to 40 frames of the test sequence *Claire* (CIF, 10 Hz). The experimental results show a reliability percentage of 85% with an average position error of 2.252 ± 0.861 pixels for the mouth corners and outer corners of the eyes, an average position error of 3.655 ± 1.854 pixels for the nose corners, and an average position error of 1.097 ± 0.744 pixels for the chin and cheek borders.

7. ACKNOWLEDGMENTS

The authors are grateful to Prof. H.-G. Musmann for encouraging this work.

8. REFERENCES

[1] K. Aizawa and T. Huang, "Model-based coding: Advanced video coding techniques for very low bit-rate applications," *Proceedings of the IEEE*, vol. 83, pp. 259-271, February 1995.

[2] R. Forchheimer, O. Fahlander, and T. Kronander, "A semantic approach to the transmission of face images," in *Picture Coding Symposium*, Cesson-Sevigne, France, July 1984, paper 10.5.

[3] H. Li, A. Lundmark, and R. Forchheimer, "Image sequence coding at very low bit-rates: A review," *IEEE Transactions on Image Processing*, vol. 3, pp. 589-609, September 1994.

[4] H. Musmann, "A layered coding system for very low bit-rate video coding," *Signal Processing: Image Communications*, vol. 7, pp. 267-278, November 1995.

[5] D. Pearson, "Developments in model-based video coding," *Proceedings IEEE*, vol. 83, pp. 892-906, June 1995.

[6] W.J. Welsh, S. Searby, and J.B. Waite, "Model-based image coding," *British Telecom Technology Journal*, vol. 8, no. 3, pp. 94-106, July 1990.

[7] G. Martinez, "Maximum-likelihood motion estimation of a human face," in *IEEE International Conference on Multimedia and Expo*, Baltimore, USA, 2003, pp. 577-580.

[8] G. Martinez, "Improving the speed of covergency of a maximum-likelihood motion estimation algorithm of a human face," in *IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, 2004.

[9] P. Ekman and V.W. Friesen, "Facial action coding system," Palo Alto, USA, 1978, Consulting Psychologists Press Inc.

[10] J. Fischl, B. Miller, and J. Robinson, "Parameter tracking in a muscle-based analysis-synthesis coding system," in *Picture Coding Symposium*, Lausanne, Switzerland, March 1993, paper 2.3.

[11] K. Aizawa, H. Harashima, and T. Saito, "Model-based analysis-synthesis image coding (mbasic) system for a person's face," *Signal Processing: Image Communicaction*, vol. 1, no. 2, pp. 139-152, October 1989.

[12] C.S. Choi, K. Aizawa, H. Harashima, and T. Takebe, "Analysis and synthesis of facial image sequences in model-based image coding," *IEEE Transactions on Circuit, Systems and Video Technology*, vol. 4, pp. 257-275, June 1994.

[13] I. Essa and A. Pentland, "Coding, analysis, interpretation and recognition of facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 757-763, July 1997.

[14] H. Li, P. Roivainen, and R. Forchheimer, "3d motion estimation in model-based facial image coding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 545-556, June 1993.

[15] M. Hess and G. Martinez, "Automatic adaptation of a human face model for model-based coding," in *Picture Coding Symposium*, San Francisco, California, December 2004.

[16] X. Zhu, J. Fan, and A.K. Elmagarmid, "Towards facial feature extraction and verification for omni-face detection in video images," in *IEEE International Conference on Image Processing*, Rochester, USA, September 2002.

[17] M. Kampmann, "Segmentation of a head into face, ears, neck and hair for knowledge-based analysis-synthesis coding of videophone sequences," in *IEEE International Conference on Image Processing*, Chicago, USA, October 1998.

[18] A. Jacquin and A. Eleftheriadis, "Automatic location tracking of faces and facial features in video sequences," *International Workshop on Automatic Face and Gesture Recognition*, pp. 142-147, 1995.

[19] K. Sobottka and I. Pitas, "Looking for faces and facial features in color images," *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, 1996.

[20] M. Kampmann and 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, pp. 201-220, March 1997.

[21] J. Ostermann and M. Kampmann, "Automatic adaptation of a facial mask in an analysis-synthesis coder based on moving flexible 3d objects," in *International Workshop on Coding Techniques for Very Low Bit-Rate Video Coding*, Colchester, United Kingdom, April 1994, No. 2.4.

[22] M. Wollborn, M. Kampmann, and R. Mech, "Content-based coding of videophone sequences using automatic face detection," in *Picture Coding Symposium*, Berlin, Germany, September 1997.

[23] M. Kampmann and G. Martinez, "Automatic adaptation of face models in videophone sequences with more than one person," in *International Workshop on Coding Techniques for Very Low Bit-Rate Video Coding*, Linkoping, Sweden, July 1997.

[24] L. Zhang, "Tracking a face in a knowledge-based analysis-synthesis coder," in *International Workshop on Coding Techniques for Very Low Bit-Rate Video Coding*, Tokyo, Japan, November 1995, No. 1.6.

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

[26] M. Kampmann and L. Zhang, "Estimation of eye, eyebrow and nose features in videophone sequences," in *International Workshop on Very Low Bit-Rate Video Coding*, Urbana, USA, October 1998.

[27] L. Zhang, "Automatic adaptation of a face model using action units for semantic coding of videophone sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 6, pp. 781-795, October 1998.

[28] L. Zhang, "Estimation of the mouth features using deformable templates," in *IEEE International Conference on Image Processing*, Santa Barbara, USA, October 1997.

[29] M. Kampmann, "Estimation of the chin and cheek contours for precise face model adaptation," in *IEEE International Conference on Image Processing*, Santa Barbara, USA, October 1997.

[30] M. Kampmann, "Automatic 3d face model adaptation for model-based coding of videophone sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 3, March 2002.

[31] S. Smith and J. Brady, "Susan - a new approach to low level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45-78, 1997.