

AUTOMATIC EXTRACTION OF FACIAL FEATURE POINTS FOR MPEG4 VIDEOPHONE APPLICATIONS

Nikolaos Sarris, Panagiotis Karagiannis and Michael G. Strintzis
Information Processing Laboratory, Electrical and Computer Engineering Department
Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

ABSTRACT

This paper addresses the problem of automatically extracting a set of characteristic feature points from images containing a front view of a human face. These feature points include edge-points of the basic characteristics of a face and are extremely useful for the adaptation of 3D head models and the extraction of Facial Animation Parameters of the Synthetic/Natural Hybrid Coding (SNHC) part of the MPEG4 image coding standard. The proposed method is based on: (a) an initial segmentation of the chrominance image to extract a rough area of the face, (b) post-processing of the extracted area by standard dilation and erosion operators, (c) geometric region splitting within the facial area and (d) detection of the features within their corresponding regions based on their geometrical characteristics.

INTRODUCTION

The segmentation of an image in general and particularly the detection and extraction of facial features are common problems in many image processing applications such as facial recognition, tracking and modeling. In the special case of videoconferencing type applications the detection of the face becomes a simpler task as the programmer is assisted by the knowledge of the scene content.

Acknowledging the individuality and usefulness of such applications the recent image coding standard MPEG4 has provided the framework for efficient coding of the human head and its expressions by means of two specific parameter sets: The Facial Definition Parameters (FDPs) and the Facial Animation Parameters (FAPs). The first describes a set of characteristic feature points on the face while the second provides a set of facial deformation parameters which have been found to be capable of describing any human expression. Having detected the positions of these feature points and calculated the values of the animation parameters, MPEG4 provides a model based coding scheme, which has shown in experimental systems to deliver videoconferencing scenes at bandwidths

as low as 10kbps [1]. However, the standard does not provide any means of performing the necessary operations in an image frame in order to detect the position of these feature points and calculate the values of the animation parameters.

In general face detection applications, techniques based on Neural Networks (NN) [4][5], Principal Component Analysis (PCA) [3], or analysis of color distribution [8], have proved to be efficient and reliable. General NN and PCA methods are based on the degree of correlation of the given image with a type of template representation of a face. In this way they manage to detect the presence and position of a face and sometimes even recognize its identity by comparing to a number of images in a given database. More complicated techniques that are based on deformable contours [7], or use of deformable 2D and 3D models [9][10] have also been reported to accurately detect the position and shape of a face. These are based on an initial contour approximation of a face, (either 2D or 3D), which is deformed by the application of specific forces until the desired fitting to the given face has been achieved

In the current application however, the extraction of facial features is necessary for the adaptation of a generic face model. These features consist of the exact locations of particular points, as outer left eye (which means the outermost/leftmost point of the left eye), inner left eye, leftmost mouth, etc. Therefore, a result in far greater detail is sought than that of detecting a rectangular, ellipsoid area or even the irregular contour containing the face.

SYSTEM DESCRIPTION

This work proposes a method for the automatic extraction of facial feature points by locating the face in the image, partitioning the face in three areas and locating the exact positions of these feature points within these areas based on the knowledge of the geometrical characteristics of the face.

The locating of the facial area (face and naked part of the neck) is based on segmenting the image using a skin color

map followed by post-processing operations to eliminate falsely detected objects, as described in [2]. A connected component-labeling algorithm is employed at the end to reject all but the largest region, as we are certain of the presence of only one face in the scene.

Having located the facial area, the part of the neck is detected and eliminated by searching for the first darkest rectangular area below the present center of the facial area. The edge of the lower chin contour is then extracted by following the most similar in luminance pixels around the detected center of the chin. Geometric methods based on moments are then employed to locate the center of gravity of the remaining face as well as its orientation axes. This information is then utilized to split the face in three regions and detect each feature in the correct region, while requiring each feature's dimensions to adhere to an expected symmetry within the facial area.

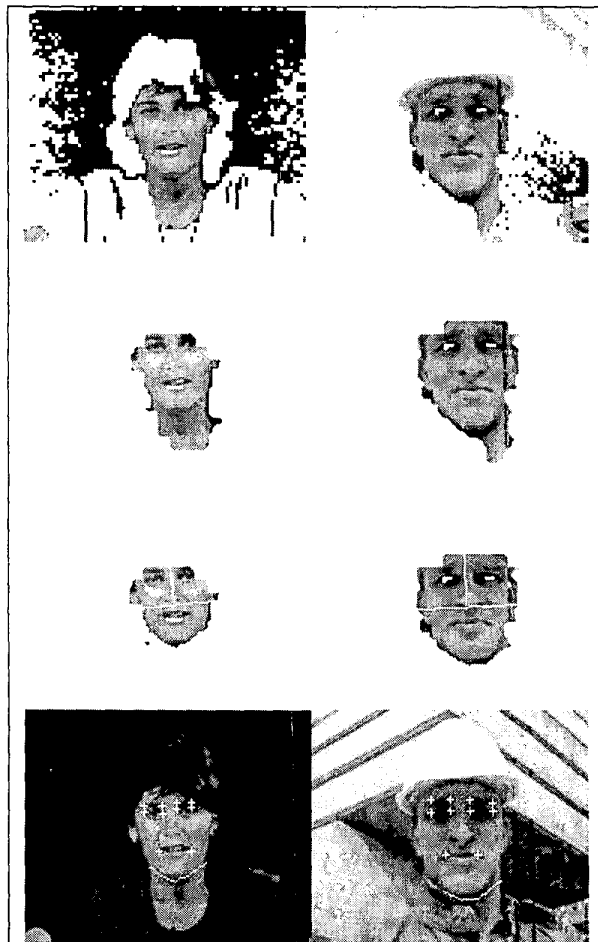


Figure 1: (a) the thresholded image (c) the face region after post-processing operations, (d) the region splitting having located the chin and (e) the detected feature points and chin contour

Example results for the common test images Miss America and Foreman are shown in Figure 1.

ACKNOWLEDGEMENTS

This work was supported by the Greek Secretariat of Research and Technology, AMEA project 'LipTelephone: A Videophone for the Deaf'

REFERENCES

- [1] G. A. Abrantes, F. Pereira, "MPEG-4 Facial Animation Technology: Survey, Implementation and Results", IEEE Transactions on CSVT, Vol. 9, No. 2, March 1999.
- [2] D. Chai, K. N. Ngan, "Face Segmentation using Skin Color Map in Videophone Applications", IEEE Transactions on CSVT, Vol. 9, No. 4, June 1999
- [3] Craw I, Costen N, Kato T, Akamatsu A, "How should we represent Faces for Automatic Recognition?", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 8, August 1999
- [4] Lawrence S, Giles L, Tsoi A, Back A, "Face Recognition: A Convolutional Neural Network Approach", IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition
- [5] Sarris Nikolaos, Tzanetos Georgios and Michael G. Strintzis: "Three Dimensional Model Adaptation and Tracking of a Human Face", 4th European Conference on Multimedia Applications, Services and Techniques, ECMAST'99 Madrid (Spain), 26-28 May 1999.
- [6] Turk M, Pentland A, "Eigenfaces for Recognition", Journal of Cognitive Science, 3:71-86, 1991
- [7] Cootes T. F., Di Mauro E. C., Taylor C. J., Lanitis A., "Flexible 3D Models from Uncalibrated Cameras", in Proc. British Machine Vision Conference, 1995, BMVA Press, pp.147-156
- [8] Sobottka K., Pittas I., "A novel method for automatic face segmentation, facial feature extraction and tracking", Signal Processing: Image Communication 12 (1998), pp 263-281
- [9] Douglas De Carlo, Dimitris Metaxas, "The integration of optical flow and deformable models with applications to human face shape and motion estimation", In Proc. CVPR '96, pp.231-238
- [10] Zhang Liang, "Tracking a face for knowledge-based coding of videophone sequences", Signal Processing: Image Communication 10 (1997), pp 93-114