Facial Feature Segmentation from Frontal View Images

Ferran Marqués and Carles Sobrevals
Dept. Teoria del Senyal i Comunicacions, Universitat Politécnica de Catalunya
UPC-Campus Nord, C/ Jordi Girona, 1-3, 08034 Barcelona
e-mail: ferran@gps.tsc.upc.es

ABSTRACT

This paper deals with the extraction of facial features from images presenting frontal views of human faces. The proposed technique combines color and structural characteristics of facial features and human faces. The technique relies on an initial generic segmentation based on color information. Regions allow a more robust estimation of the facial features location. Mouth region candidates are first selected based on color and shape information. From these mouth candidates, and applying geometrical information, eyes and eyebrows region candidates are defined. All sets of candidates are jointly analyzed and the most likely set, from a structural viewpoint, is selected. The technique has been tested with a large set of images, both from fixed frontal view databases (XM2VTS) as well as from generic ones (MPEG4 and MPEG7 databases).

1 INTRODUCTION

Currently, there are several applications in image and video analysis that require the extraction and tracking of facial human features (e.g.: expression recognition or facial animation) [1, 2]. Several techniques have been proposed in the literature to solve this problem relying on model based approaches or on color and structure information [3].

The proposed facial feature extraction technique tries to increase the robustness color and structure based techniques avoid working at pixel level. Therefore, as first step, it segments the image into homogeneous regions. Image analysis using regions allows a more robust estimation of the location of facial features. The segmentation is computed using color information to achieve more accurate contours. Using a color homogeneous partition, mouth region candidates are first selected based on color and shape information. Based on these mouth candidates, and applying geometrical information, eyes and eyebrows (when possible) region candidates are defined. Finally, all sets of candidates are jointly analyzed and the most likely set, from a structural viewpoint, is selected.

The paper is structured as follows. Section 2 deals with the generic segmentation technique that has been used in this work. In Section 3, the facial feature extraction technique is discussed. Section 4 presents several results obtained applying the proposed technique to images from frontal face databases and generic ones. In addition, the limitations of the technique are discussed. Finally, in Section 5, some conclusions are driven and our current work in this area is commented.

2 GENERIC SEGMENTATION

In this work, we are not trying to perform a specific segmentation for the application of facial feature detection. On the contrary, we want to be able to extract facial features from a generic segmentation. Of course, the technique will be limited by the presence of such features in the resultant partition. Facial features will be obtained as single regions or the union of regions belonging to this initial partition. Thus, this partition has to ensure that it represents, as correctly as possible, the facial feature contours. Figure 1 shows an example of this type of initial partition.

Figure 1. Original image and example of initial partition where regions forming the facial features are highlighted

We have analyzed various segmentation techniques using different segmentation criteria. Results are quite similar when using a segmentation criterion that takes into account color information. In our case, we use a region growing approach [4] based on an Euclidean weighted distance in the (y, u, v) color space, where more relevance is given to the luminance component:

$$d(r_1,r_2) = \sqrt{\alpha\left(\bar{y}_1 - \bar{y}_2\right)^2 + \frac{1-\alpha}{2}\left(\left(\bar{u}_1 - \bar{u}_2\right)^2 + \left(\bar{v}_1 - \bar{v}_2\right)^2\right)}$$

The merging order between pairs of neighbor regions is based on the previous distance. Regions (r_1, r_2) are compared using the mean values of their (y, u, v) components. The weight α is tipycally set to 0.5.

3 FEATURE EXTRACTION

As previously commented, the selection of those regions that represent the facial features is based on color, shape and structural information. The facial structure (relative position of the different facial features and their aspect ratios) is adopted from the CANDIDE model [5].

3.1 Mouth candidates extraction

Initially, mouth candidates are obtained by selecting those regions representing a local maximum of the v color component. Candidate regions are ordered by their aspect ratio; that is, the longest region is ranked first. Mouth regions are completed starting by the best ranked candidate. A rectangular searching area is established around the region candidate and its neighbor regions are merged to it if their v color component values are close and the region resulting from the union improves the aspect ratio.

3.2 Eyes candidates selection

Eye candidates are usually composed of several regions from which, at least one, corresponds to a local minimum of the v color component or a local maximum of the u color component. All regions presenting any of these characteristics are selected as eye candidates. This way, eye and eyebrow regions (when present) are initially selected as eye candidates.

In order to increase the robustness againts poses as well as partial mouth initial detections, the positions of the mouth candidates introduce a very soft constraint in the selection of eye candidates. This way, eye candidates are only forced to be above mouth candidates.

3.3 Eye candidate pairs creation

Starting from the mouth candidate location, pairs of eye regions are created imposing that they should be at the same distance from the center of mass of the mouth region. Furthermore, it is impossed that the distance between both eyes should be related to the distance between the mouth and the eyes. The relation between both distances is given by the CANDIDE model (1,27) and a small tolerance has been included to allow some deviation. Such a tolerance is higher in the case of the distance between eyes since it varys more among different people.

Therefore, for each eye candidate two circles are computed, each one representing one of the previous distances, and the associated eye candidate is sought in the intersetion of both circles. Typically, this step selects two different pairs of candidates: the eyes and the eyebrows. In the case of presence of glasses or some specific problems in the skin, other erronous pairs may appear.

3.4 Mouth and eyes validation

The set of pair of eye candidates for a given mouth is analyzed and compared here. If the previous step has just provided with a single pair of candidates, they are assumed to represent the eyes. In the case of obtaining more than a single pair, the possibility of having the eyes and the eyebrows is analyzed. If two pairs of candidates are very close and one above the other, they are assumed to represent the eyes and eyebrows. In the case of having several candidates and none of them can be considered to be eyebrows, we select the pair leading to a ratio between the distances eye-eye and eye-mouth closer to the CANDIDE model ratio.

3.5 Mouth completion

Once the eyes have been detected, the mouth region can be completed if it was, at the first stage, only partially detected. With the detection of both eyes, a medial axis can be estimated and the mouth is assumed to be symmetrical with respect to it. Therefore, new regions are merged to the mouth candidate, allowing to include non-connected components. This way, both lips can be detected even in the case of having the mouth openned.

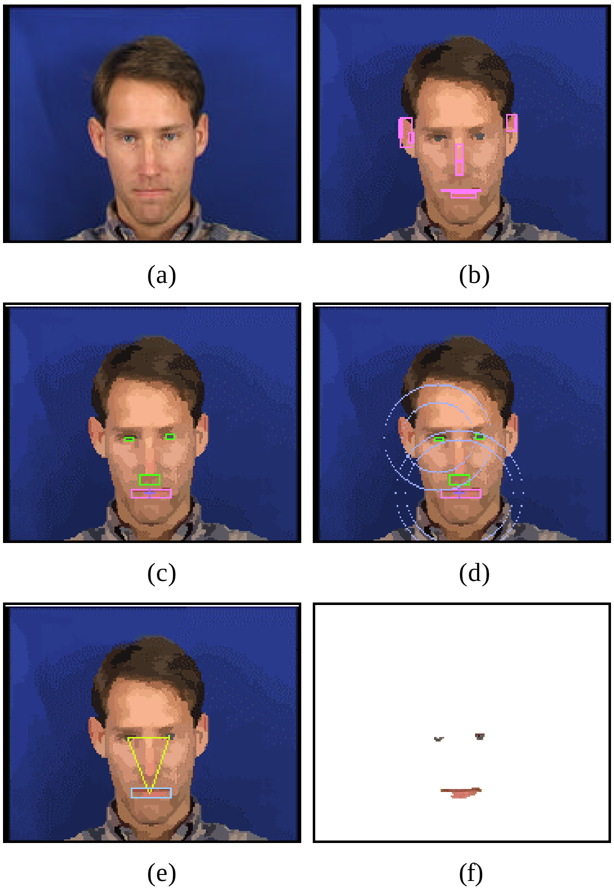

Figure 2. Example of the complete process: (a) Original Image, (b) Mouth candidates, (c) Eye candidates for a given mouth candidate, (d) Creation of pairs of eye candidates, (e) Validated regions, (f) Final extracted facial features.

In Figure 2, a complete example of the feature extraction process is presented. Note that, given that the partition does not present separate regions representing the eyebrows, they have not been detected in this case. Moreover, the final mouth completion step does not improve the mouth extraction since the growing of the initial mouth candidate already provides with a good mouth representation.

4 RESULTS

The previous technique has been applied to the first shot of the XM2VTS database [6] of frontal views, which contains one frontal view per person (289 different people). Table 1 summarizes the global results obtained. As it can be seen, the algorithm performs correctly in a 85,5% of the cases.

Description	# Images	Percentage
Complete database	289	100 %
Incomplete partition	6	2.0 %
Missing one eye candidate	7	2.5 %
Wrong pair selection	16	5.5 %
Wrong mouth detection	3	1.0 %
No features found	10	3.5 %
Mouth and eyes correct	247	85.5 %

Table 1. Global results on XM2VTS frontal view database.

It has to be highlighted that, from the 42 persons that are not correctly analyzed, 26 wear glasses (62% of the errors). Glasses disturb the correct segmentation of eyes since they yield fuzzy regions that, due to reflections, may not have to correct color components and, therefore, they are not selected as candidates (7). In some extreme cases (6), eyes are not even present in the partition. Moreover, frames may produce wrong candidates that due to their symmetry and location can be selected as correct ones. They correspond to the largest part of the 16 wrong selections of pair of regions as eyes. There are a few cases (3) where the mouth is wrongly detected, usually due to the presence of reddish clothes. Finally, there are some images (10) in which the algorithm fails to detect any feature because of various different reasons (e.g.: bearded men).

Figure 3 shows some results obtained with the XM2VTS database. In them, we have tried to illustrate the behavior of the algorithm. In the first one, all facial features (mouth, eyes and eyebrows) are correctly obtained. Furthermore, nostrils are extracted. Nostrils, when present in the partition, are easily extrated since they are represented as dark regions whose positions can be robustly determined. In the second case, an example of a person wearing a moustach and glasses is presented. Note that the eyes are correctly obtained and so is the mouth. However, the mouth region is completed in the last step, once the eyes have been validated. The third case has been selected in order to show how the algorithm performs when analyzing a blonde person. As it can be seen, all facial features are correctly extracted. The fourth case presents an example of a person with a pose and wearing glasses. As it can be seen, the frame completely hides the eyebrows whereas, luckly, light reflections in the glasses do not overlap the eyes. In the last example, the correct feature detection in the case of a bearded man is presented.

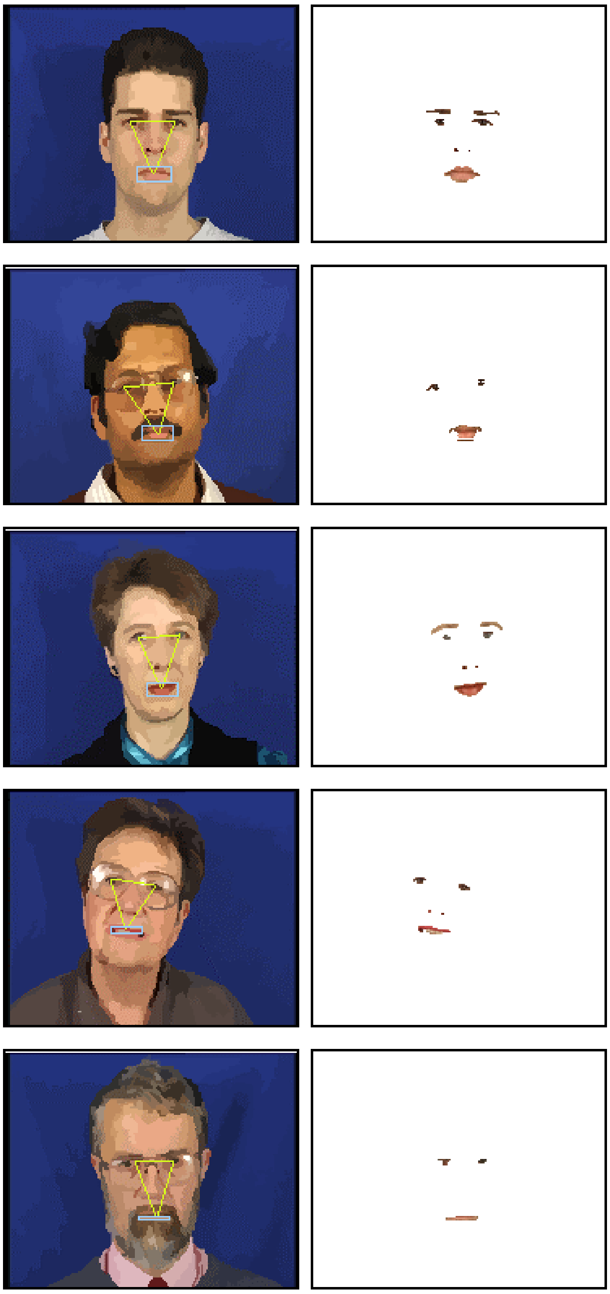

Figure 3. Five examples of facial feature extraction from the XM2VTS database.

In Figure 4, some additional results are presented. In this case, the previous technique is applied to images that have been captured in normal conditions and contain frontal (or quasi frontal) views of persons. In these examples one can note the robustness of the technique. Even in the case of complex backgrounds, presence of several reddish regions or of a person presenting a pose, the algorithm is able to correctly extract those features that appear in the partition.

Figure 4. Five examples of facial feature extraction in generic images containing a human face.

4 CONCLUSIONS

This paper has presented a new approach for obtaining facial features from frontal views of human faces. The quality and robustness of the proposed technique has been assessed using images from frontal human view databases as well as generic images containing frontal human views. The average computational time is 40 miliseconds on a Pentium III. The technique, which is based on color and structural information, performs all the analysis after segmenting the original image. This initial segmentation is carried out in order to make more robust the estimation of the facial feature locations. However, the use of this initial segmentation introduces a constraint in the algorithm: those features that do not appear as regions in the segmentation cannot be detected.

The current work tries to improve the robustness of the tecniche as well as to extend it to more complex cases: first, we are analyzing the way to improve the initial partition when a feature is not detected. The idea is to refine the partition in the area where the analysis prompts that a possible facial feature is missing. Second, for those cases where the scenario is very complex, we are including a face detection step [7] that will restrict the area of search. Finally, we are extending the complete technique to cope with profile human views.

ACKNOWLEDGMENTS

This work has been partly supported by the CEE project InterFace (IST-1999-10036) and by the grant CICYT TIC2001-0996 of the Spanish Government.

REFERENCES

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

[2] S. Morishima, "Face analysis and synthesis for duplication expression and impression", *IEEE Signal Processing Magazine*, V. 18, No. 3, pp. 26-34, May 2001.

[3] M. Pantic, L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art", IEEE Trans. on PAMI, Vol. 22, N. 12, pp. 1424-1445, December 2000.

[4] P. Salembier, L. Garrido "Binary partition tree as an efficient representation for image processing and information retrieval", *IEEE Trans. on Image Processing*, 9(4):561-576, April 2000.

[5] R. Forchheimer, T. Kronander, "Image coding – from waveforms to animation" *IEEE Trans. on ASSP*, Vol. 37, N. 12, Dec. 1989..

[6] http://www.ee.surrey.ac.uk/research/VSSP/xm2vtsdb.

[7] F. Marqués, V. Vilaplana, "Face segmentation and tracking based on connected operators and partition projection", *Pattern Recognition*, to be published in 2002.