DETECTION OF FACIAL CHARACTERISTICS BASED ON EDGE INFORMATION

Stylianos Asteriadis, Nikolaos Nikolaidis, Ioannis Pitas
Department of Informatics, Aristotle University of Thessaloniki, Box 451 Thessaloniki, GR-54124, Greece

Montse Pardas
Dept. Teoria del Senyal i Comunicacions, Universitat Politécnica de Catalunya, Barcelona, Spain

Keywords: Facial features detection, distance map, ellipse fitting.

Abstract: In this paper, a novel method for eye and mouth detection and eye center and mouth corner localization, based on geometrical information is presented. First, a face detector is applied to detect the facial region, and the edge map of this region is extracted. A vector pointing to the closest edge pixel is then assigned to every pixel. x and y components of these vectors are used to detect the eyes and mouth. For eye center localization, intensity information is used, after removing unwanted effects, such as light reflections. For the detection of the mouth corners, the hue channel of the lip area is used. The proposed method can work efficiently on low-resolution images and has been tested on the XM2VTS database with very good results.

1 INTRODUCTION

In recent bibliography, numerous papers have been published in the area of facial feature localization, since this task is essential for a number of important applications like face recognition, human-computer interaction, facial expression recognition, surveillance, etc.

In (Cristinacce et al., 2004) a multi-stage approach is used to locate features on a face. First, the face is detected using the boosted cascaded classifier algorithm by Viola and Jones (Viola and Jones, 2001). The same classifier is trained using facial feature patches to detect facial features. A novel shape constraint, the Pairwise Reinforcement of Feature Responses (PRFR) is used to improve the localization accuracy of the detected features. In (Jesorsky et al., 2001) a three stage technique is used for eye center localization. The Hausdorff distance between edges of the image and an edge model of the face is used to detect the face area. At the second stage, the Hausdorff distance between the image edges and a more refined model of the area around the eyes is used for more accurate localization of the upper area of the head. Finally, a Multi-Layer Perceptron (MLP) is used for finding the exact pupil locations. In (Zhou and Geng, 2004) the authors use Generalized Projection Functions (GPF) to locate the eye centers in an eye area found using the algorithm proposed in (Wu and Zhou, 2003). The type of functions used here are a linear combination of functions which consider the mean of intensities and functions which consider the intensity variance along rows and columns.

A technique for eyes and mouth detection and eyes center and mouth corners localization is proposed in this paper. After accurate face detection using an ellipse fitting algorithm (Salerno et al., 2004), the detected face is normalized to certain dimensions and a vector field is created by assigning to each pixel a vector pointing to the closest edge. The eyes and mouth regions are detected by finding regions inside the face, whose vector fields resemble the vector fields of eye and mouth templates extracted from sample eye and lip images. Intensity and color information is then used within the detected eye and mouth regions in order to accurately localize the eye centers and the mouth corners. Our technique has been tested on the XM2VTS database with very promising results. Comparisons with other state of the art methods verify that out method achieves superior performance.

The structure of the paper is as follows. Section 2 describes the steps followed to detect a face in an image. In section 3 the method used to locate eye

Asteriadis S., Nikolaidis N., Pitas I. and Pardas M. (2007).
DETECTION OF FACIAL CHARACTERISTICS BASED ON EDGE INFORMATION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 247-252

and mouth areas on a face image is described. Section 4 details the steps used to localize the eye centers and the mouth corners. Section 5 describes the experimental evaluation procedure (database, distance metrics, etc). In section 6 results on eye and mouth detection are presented, and the proposed method is compared to other approaches in the literature. Conclusions follow.

2 FACE DETECTION

Prior to eye and mouth region detection, face detection is applied on the face images. The face is detected using the Boosted Cascade method, described in (Viola and Jones, 2001). The output of this method is usually the face region with some background. Furthermore, the position of the face is often not centered in the detected sub-image, as it is shown in the first row of Figure 1. Since the detection of the eyes and mouth will be done on detected face regions of a predefined size, it is very important to have a very accurate face detection. Consequently, a technique to postprocess the results of the face detector is used.

More specifically, a technique that compares the shape of a face with that of an ellipse is used. This technique is based on the work reported in (Salerno et al., 2004). According to this technique, the distance map of the face area found at the first step is extracted. Here, the distance map is calculated from the binary edge map of the area. An ellipsis scans the distance map and a score that is the average of all distance map values on the ellipse contour e, is evaluated.

$$score = \frac{1}{|e|} \sum_{(x,y)\in e} D(x,y) \tag{1}$$

where D is the distance map of the region found by the Boosted Cascade algorithm and $|e|$ denotes the number of the pixels covered by the ellipse contour.

This score is calculated for various scale and shape transformations of the ellipse. The transformation which gives the best score is considered as the one that corresponds to the ellipse that best describes the exact face contour.

The right and left boundaries of the ellipse are considered as the new lateral boundaries of the face region. Examples of the ellipse fitting procedure are shown in Figure 1.

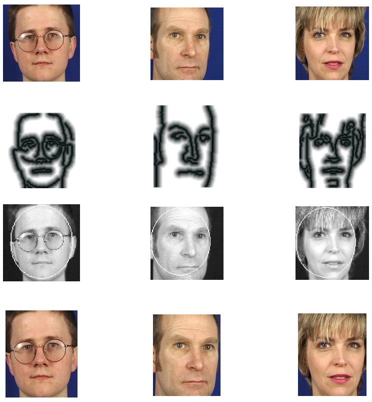
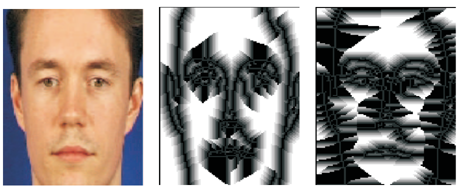


Figure 1: Face detection refinement procedure.

3 EYE AND MOUTH REGION DETECTION

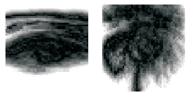
3.1 Eye Region Detection

The proposed eye region detection method can be outlined as follows: The face area found is scaled to 150x105 pixels and the Canny edge detector is applied. Then, for each pixel, the vector that points to the closest edge pixel is calculated. The x and y components of each vector are assigned to the corresponding pixel. Thus, instead of the intensity values of each pixel, we generate and use in the proposed algorithm a vector field whose dimensions are equal to those of the image, and where each pixel is characterized by the vector described above. This vector encodes for each pixel, information regarding its geometric relation with neighboring edges, and thus is relatively insensitive to intensity variations or poor lighting conditions. The vector field can be represented as two maps (images) representing the horizontal and vertical vector components for each pixel. Figure 2(a) depicts the detected face in an image, Figures 2(b) 2(c), show the horizontal and vertical component maps of the vector field of the detected face area, respectively.



(a) (b) (c)

Figure 2: (a) Detected face (b) horizontal coordinates of vector field (c) vertical coordinates of vector field.



(a) (b)

Figure 3: (a) Mean vertical component map of right eye (b) mean horizontal component map of right eye.

In order to detect the eye areas, regions R_k of size NxM within the detected face are examined and the corresponding vector fields are compared with the mean vector fields extracted from a set of right or left eye images (Figure 3).

The similarity between an image region and the templates is evaluated by using the following distance measure:

$$E_{L_2} = \sum_{i \in R_k} \|\mathbf{v}_i - \mathbf{m}_i\| \qquad (2)$$

where $\| \cdot \|$ denotes the L_2 norm. Essentially for a NxM region R_k the previous formula is the sum of the euclidean distances between vectors $\mathbf{v}_i$ of the candidate region and the corresponding $\mathbf{m}_i$ of the mean vector field of the eye we are searching for (right or left). The candidate region on the face that minimizes E_{L_2} is marked as the region of the left or right eye.

3.2 Mouth Region Detection

(a) (b) (c)

Figure 4: (a) Sample mouth region image, (b) mean vertical component map of the mouth region, (c) mean horizontal component map.

The mouth region was detected using a procedure similar to the one used for eye detection. The vector field of various candidate regions was compared to a mean mouth vector field. For the extraction of this mean vector map, mouth images, scaled to the same dimensions N_mxM_m, were used. An example of a mouth image, used for the calculation of the mean vector field and the mean horizontal and vertical component maps, can be seen in Figure 4(a). However, since lip and skin color are, in many cases, similar and since beard (when existent) might occlude or distort the lips shape, lips localization is more difficult. For this reason, an additional factor is included in eq. 2. This factor is the inverse of the number of edge pixels of the horizontal edge map evaluated within the candidate mouth area. This term was added because, due to the elongated shape of the lips, the corresponding area is characterized by a large concentration of horizontal edges. Thus, this factor helps at discriminating between mouth/non-mouth regions. The additional factor is weighted so that its mean value in the search zone is equal to the mean value of the E_{L_2} distance of the candidate mouth areas from the mean vector coordinate maps. Based on the above, the distance measure used in mouth region detection is the following:

$$E_{L_2}^{mouth} = \sum_{i \in R_k} \|\mathbf{v}_i - \mathbf{m}_i\| + \frac{w}{\sum_{i \in R_k} I_i^{horizontalEdges}}, \qquad (3)$$

where $I_i^{horizontalEdges}$ is the horizontal binary edge value for pixel i of candidate region R_k. More specifically, $I_i^{horizontalEdges}$ is one if pixel i is an edge pixel and zero otherwise.

4 LOCALIZATION OF CHARACTERISTIC POINTS

After eye and mouth areas detection, the eye centers and mouth corners are localized within the found areas using the procedures described in the following sections.

4.1 Eye Center Localization

The eye area found using the procedure described in section 3 is scaled back to the dimensions N_{eye}xM_{eye} it had in the initial image. Moreover, before eye center localization, a pre-processing step is applied. Since reflections (highlights), that affect the results in a negative way, frequently appear on the eye, a reflection removal step is implemented. This proceeds as follows: The eye area is first converted into a binary image through thresholding using the threshold selection method proposed in (Otsu, 1979). Subsequently, all the small white connected components of the resulting binary eye image are considered as highlight areas and the intensities of the pixels in the grayscale image that correspond to these areas are substituted by the average luminance of their surrounding pixels. The result is an eye area with most highlights removed.

The eye center localization is performed in three steps, each step refining the results obtained in the previous one. By inspecting the eye images used for the extraction of the mean vector maps, one can observe that the eyes reside at the lower central part of the detected eye area. Thus, the eye center is searched within an area that covers the lower 60% of the eye region and excludes the right and left parts of this region. The information in this area comes from the eye itself and not from the eyebrow or the eyeglasses.

Since, at the actual eye center position, there is significant luminance variation along the horizontal and vertical axes, the images $D_x(x,y)$ and $D_y(x,y)$ of the absolute discrete intensity derivatives along the horizontal and vertical directions are evaluated:

$$D_x(x,y) = |I(x,y) - I(x-1,y)| \quad (4)$$

$$D_y(x,y) = |I(x,y) - I(x,y-1)| \quad (5)$$

The contents of the horizontal derivative image are subsequently projected on the vertical axis and the contents of the vertical derivative image are projected on the horizontal axis. The 4 vertical and 4 horizontal lines, corresponding to the 4 largest vertical and horizontal projections (i.e., the lines crossing the strongest edges) are selected. The point whose x and y coordinates are the medians of the coordinates of the vertical and horizontal lines respectively, defines an initial estimate of the eye center (Figure 5(a)).

Using the fact that the eye center is in the middle of the largest dark area in the region, the previous result can be further refined: The darkest column (defined as the column with the lowest sum of pixel intensities) of a $0.4N_{eye}$ pixels high and $0.15M_{eye}$ pixels wide area around the initial estimate is found and its position is used to define the horizontal coordinate of the refined eye center. In a similar way, the darkest row in a $0.15N_{eye}$x$0.4M_{eye}$ area around the initial estimate is used to locate the vertical position of the eye center (Figure 5(b)).

For even more refined results, in a $0.4N_{eye}$xM_{eye} area around the point found at the previous step, the darkest $0.25N_{eye}$x$0.25M_{eye}$ region is searched for, and the eye center is considered to be located in the middle of this region. This point gives the final estimate of the eye center, as can be seen in figure 5(c).

(a) (b) (c)

Figure 5: (a)Initial estimate of eye center (b) estimate after first refinement, (c) final eye center localization.

4.2 Mouth Corner Localization

For mouth corner localization, the hue component of mouth regions can be exploited, since the hue values of the lips are distinct from those of the surrounding area. More specifically, the lip color is reddish and, thus, its hue values are concentrated around 0^o. In order to detect the mouth corners, the pixels of the hue component are classified into two classes through binarization (Otsu, 1979). The class whose mean value is closer to 0^o is declared as the lip class. Small components assigned to the lip class (while they are not lip parts) are discarded using a procedure similar to the light reflection removal procedure.

Afterwards, the actual mouth corner localization is performed by scanning the binary image and looking for the rightmost and leftmost pixels belonging to the lip class.

5 EXPERIMENTAL EVALUATION PROCEDURE

The proposed method has been tested on the XM2VTS database (Messer et al., 1999), which has been used in many facial feature detection papers. This database contains 1180 face and shoulders images. All images were taken under controlled lighting conditions and the background is uniform. The database contains ground truth data for eye centers and mouth corners.

Out of a total of 1180 images, only 3 faces failed to be detected. In cases of more than one candidate face regions in an image, the smallest sum of the distance metric (eq. 2) for the left and right eye and the distance metric (eq. 3) for the detected mouth was retained, in order for false alarms to be rejected.

For eye region detection, success or failure was declared depending on whether the ground truth for both eye centers was in the found eye regions. Mouth region detection was considered successful if both ground truth mouth corners were inside the region found. For the eye center and mouth corner localization, the correct detection rates were calculated through the following criterion, introduced in (Jesorsky et al., 2001):

$$m_2 = \frac{max(d_1, d_2)}{s} < T \quad (6)$$

In the previous formula, d_1 and d_2 are the distances between the eye centers or mouth corners ground truth and the eye centers or mouth corners found by the algorithm, and s is the distance between the two ground truth eye centers or the distance between the mouth corners. A successful detection is declared whenever m_2 is lower than threshold T.

6 EXPERIMENTAL RESULTS

Two types of results were obtained on the images described above: results regarding eye/lips region detection and results on eye center/mouth corner localization. All the results take into account the results of the face detection step and are described in the following sections.

6.1 Eye Detection and Eye Center Localization

Correct eye region detection percentages are listed in the column of Table 1 denoted as ”Eye regions”. It is obvious that the detection rates are very good both for people not wearing eyeglasses and those who do.

The column labelled ”Eye Centers” in the same Table present correct eye center localization results for threshold value T=0.25.

Furthermore, the success rates for various values of the threshold *T*, for the whole database are depicted in Figure 6. From the figure it can be observed that, even for very small thresholds T (i.e. for very strict criteria), success rates remain very high. For example, the maximum distance of the detected eye centers from the real ones does not exceed 5% (T=0.05) of the inter-ocular distance in 93.5% of the cases, which means that the algorithm can detect eye centers very accurately.

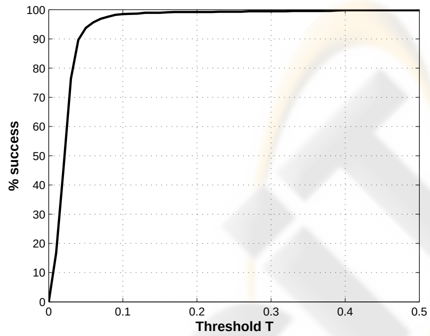


Figure 6: Eye center localization for various thresholds T.

6.2 Mouth Detection and Mouth Corner Localization

The mouth was correctly detected in 98.05% of the cases. The mouth corner localization success rates for T=0.25 is 97.6%. Figure 7 shows the success rates of mouth corner localization for various T. It is obvious that the method has very good performance in detecting the mouth and localizing its corners.

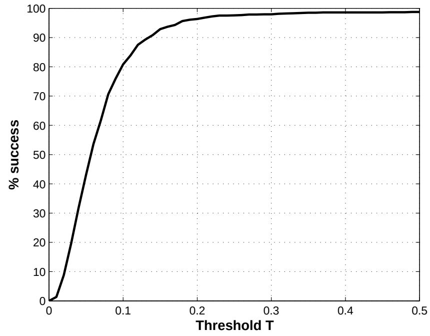


Figure 7: Mouth corner localization for various thresholds T for the entire database.

6.3 Comparison with Other Methods

The method has been compared with other existing methods, that were tested by the corresponding authors on the same database for the eye center localization task. Unfortunately, no mouth corner detection method tested on the XM2VTS database was found. For T=0.25 our method achieves an overall detection rate of 99.3%, while Jesorsky *et al* in (Jesorsky et al., 2001) achieve 98.4%. The superiority of the proposed method is much more prominent for stricter criteria, i.e. for smaller values of the threshold T: For T=0.1, both (Jesorsky et al., 2001) and (Cristinacce et al., 2004) achieve a success rate of 93%, while the proposed method localizes the eye centers successfully in 98.4% of the cases. Some results of the proposed method can be seen in Figure 8.

7 CONCLUSIONS

A novel method for facial feature detection and localization was proposed in this paper. The method utilizes the vector field that is formed by assigning to each pixel a vector pointing to the closest edge, encoding, in this way, the geometry of such regions, in order to detect eye and mouth areas. Luminance and chromatic information were exploited for accurate localization of characteristic points, namely the eye centers and mouth corners. The method proved to give very accurate results, failing only at extreme cases.

ACKNOWLEDGEMENTS

This work has been partially supported by the FP6 European Union Network of Excellence MUSCLE

Table 1: Results on the XM2VTS database.

	Eye Regions	Eye Centers for T=0.25
People without glasses	99.2%	99.6%
People with glasses	98.2%	98.7%
Total	98.85%	99.3%

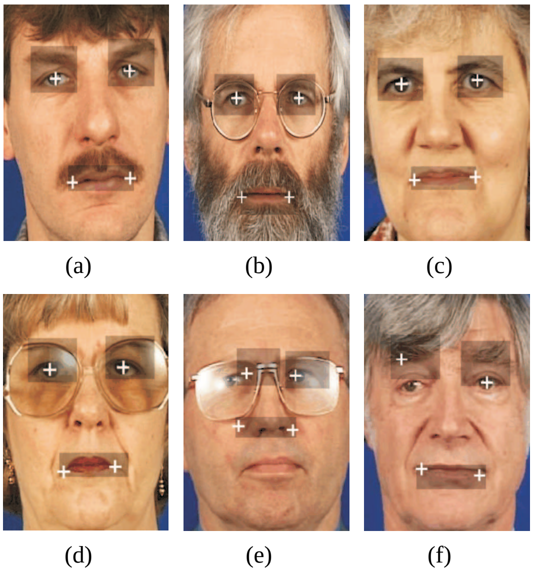


Figure 8: Some successfully (a)-(d) and some erroneously (e),(f) detected facial features.

”Multimedia Understanding Through Semantic Computation and Learning” (FP6-507752).

REFERENCES

Cristinacce, D., Cootes, T., and Scott, I. (2004). A multi-stage approach to facial feature detection. In *15th British Machine Vision Conference, 231-240*.

Jesorsky, O., Kirchberg, K. J., and Frischholz, R. W. (2001). Robust face detection using the hausdorff distance. In *3rd International Conference on Audio and Video-based Biometric Person Authentication, 90-95*.

Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. (1999). Xm2vtsdb: The extended m2vts database. In *2nd International Conference on Audio and Video-based Biometric Person Authentication, 72-77*.

Otsu, N. (1979). A threshold selection method from gray-level histograms. In *IEEE Transactions on Systems, Man, and Cybernetics, 9, No 1, 62-66*.

Salerno, O., Pardas, M., Vilaplana, V., and Marques, F. (2004). Object recognition based on binary partition trees. In *IEEE Int. Conference on Image Processing, 929-932*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Vision and Pattern Recognition, 1, 511-518*.

Wu, J. and Zhou, Z. (2003). Efficient face candidates selector for face detection. In *Pattern Recognition, 36, No 5, 1175-1186*.

Zhou, Z. and Geng, X. (2004). Projection functions for eye detection. In *Pattern Recognition, 37, No 5, 1049-1056*.