The ENCARA System for Face Detection and Normalization

M. Castrillón, O. Déniz, M. Hernández

IUSIANI
Edificio Central del Parque Científico-Tecnológico
Campus Universitario de Tafira
Universidad de Las Palmas de Gran Canaria
35017 Las Palmas - Spain

Abstract

Facial image processing is becoming widespread in human-computer applications, despite its complexity. High-level processes such as face recognition or gender determination rely on low-level routines that must effectively detect and normalize the faces that appear in the input image. In this paper, a face detection and normalization system is described. The approach taken is based on a cascade of fast, weak classifiers that together try to determine whether a frontal face is present in the image. The system is also able to obtain the position of facial features, such as mouth and eyes, and it operates in real-time. Comprehensive experiments carried out with real video sequences show that the system is faster than other approaches and effective in detecting frontal faces.

Keywords: computer vision, face detection, facial analysis, face recognition.

1   Introduction

Faces are the center of human-human communication, and have been object of analysis for centuries. It is evident, that the face conveys to humans such a wealth of social signals, and humans are expert at reading them. They tell us who is the person in front of us or help us to guess features that are interesting for social interaction such as gender, age, expression and more. That ability allows us to react differently with a person based on the information extracted visually from his/her face. For these and other reasons, computer-based facial analysis is becoming widespread, covering applications such as identity recognition, gender determination, facial expression detection, etc.

This work focus on real time face detection. Face detection must be a necessary preprocessing step in any automatic face recognition [1] or face expression analysis system [2]. However, the face detection problem commonly has not been considered in depth, being treated as a previous step in a more categorical system. Thus, many face recognition systems in the literature assume that a face has already been detected before performing the matching with the learned models

[1]. This is evidenced by the fact that face detection surveys are very recent in the Computer Vision community [4,10].

The standard face detection problem given an arbitrary image can be defined as: *to determine any face -if any- in the image returning the location and extent of each* [4,10]. The whole procedure must perform in a robust manner for illumination, scale and orientation changes in the subject. It should be noticed that trying to build a system as robust as possible, i.e., detecting any possible facial pose at any size and under any condition, seems to be an extremely hard and certainly not trivial problem. As an example, a surveillance system can not expect that people show their faces clearly. Such a system must work continuously and should keep on looking at the person until he or she offers a good opportunity for the system to get a frontal view, or make use of multimodal information with an extended focus. Thus, robustness is a main aspect that must be taken into account by any system.

Face detection methods can be classified according to different criteria, and certainly some methods overlap different categories under any classification [4,10]. In this paper, these techniques are classified into two main families according to the information used to model faces:

- Pattern based (Implicit): These approaches work mainly on still gray images. They work searching a pattern at every position of the input image, applying the same procedure to the whole image.
- Knowledge based (Explicit): These approaches reduce processing costs taking into account face knowledge explicitly, exploiting and combining cues such as color, motion, face and facial features geometry, facial features appearance and temporal coherence for sequences.

The system presented here can be related to both categories, as it makes use of both implicit and explicit knowledge. The paper is organized as follows: in Section 2 the proposed system is described, in Section 3 results of experiments carried out with it are analyzed. Finally, in Section 4 the main conclusions of the work are outlined, as well as directions for future research.

2   The ENCARA system

Some facts have been considered during the development of the ENCARA face detection system. They can be summarized as follows:

1. ENCARA is designed to detect frontal faces in video streams. ENCARA is developed for providing fast performance in human-computer interaction applications where just soft recognition would be necessary.
2. ENCARA makes use only of visual information provided by a single camera. Its performance must be good enough using standard webcams.
3. ENCARA makes use of explicit and implicit knowledge.
4. ENCARA is based on a cascade hypothesis/verification classification schema.
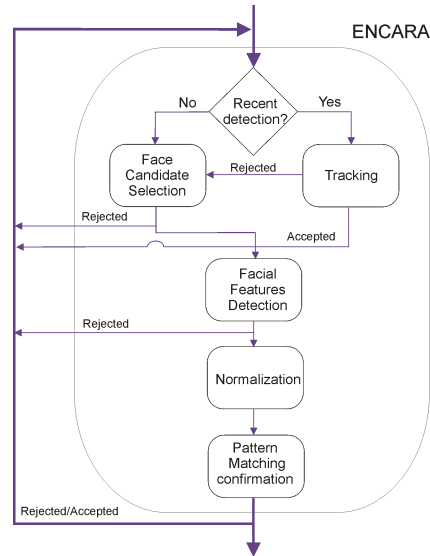
ENCARA

No   Recent
      detection?   Yes

Face
Candidate   Rejected   Tracking
Selection

Rejected

Accepted

Facial
Features
Detection

Rejected

Normalization

Pattern
Matching
confirmation

Rejected/Accepted

Figure 1. ENCARA general modules

5. Finally, the system is open to integrate new modules or modify the existing ones. This feature allows the system to incorporate improvements.

The process launches an initial face hypothesis on selected areas in the image. These areas present some kind of evidence that make them valid to assume that hypothesis. Later, the problem is tackled making use of multiple simple techniques applied opportunistically in a cascade approach in order to confirm/reject the initial frontal face hypothesis. In the first case, the module results are passed to the following module. In the second, the area is rejected. Those techniques are combined and coordinated with temporal information extracted from a video stream to improve performance. They are based on contextual knowledge about face geometry, appearance and temporal coherence.

ENCARA is described in terms of the main modules described in Figure 1. For each module the literature offers many valid techniques but in this implementation, the attention was paid to process at frame rate. The process is as follows:

1. *Tracking (M0)*: ENCARA processes a video stream, if there was a recent frontal face detection, the system tries first to track facial features instead of detecting them again. If any of these tests is passed, the candidate is accepted as frontal. In any other case, the process continues.
   (a) *Last eye and mouth search:* The last detected patterns are searched in new frames.
   (b) *Test with previous:* If the tracked positions are similar to the one in previous frame, ENCARA applies the appearance test.
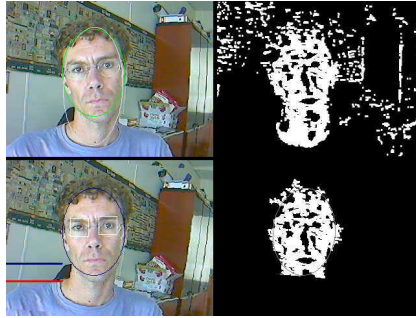
Figure 2. Example of resulting blob after neck elimination.

(c) *Majority test:* If the test with the previous frame is not passed, an extra test is performed to check if most patterns corresponding to facial features have not been lost and are located close to the previous position.

2. *Face Candidate Selection, M1:* The following steps are carried out to select areas of interest:

   (a) *Color Blob Detection:* Once the normalized red and green [9] image has been calculated, a simple schema based on defining a rectangular discrimination area on that color space is employed for skin color classification. Dilation is applied to the resulting blob image using a 3×3 element.

3. *Facial Features Detection, M2:* ENCARA searches facial features:

   (a) *Ellipse Approximation:* Major blobs detected as skin are fitted to a general ellipse using the technique described in [8] that returns the area, orientation and axis lengths of the ellipse in pixel units.

   (b) *Refusing Ellipses:* Before going further, some face candidates are rejected based on the dimensions of the ellipse detected and the axis.

   (c) *Rotation:* For this problem, it has been considered that a face can be rotated from its vertical position no more than 90 degrees, i.e., the hair is always over the chin. The orientation obtained from the ellipse fitted is employed for rotating the source image in order to get a face image where both eyes should lie on a horizontal line.

   (d) *Neck Elimination:* The quality of ellipse fitting mechanism is critical for the procedure. Clothes and hair styles affect the shape of the blob. If all these pixels that are not face such as neck and shoulders are not avoided, the rest of the process will be influenced by a bad estimation of the face area. This blob shape uncertainty will later affect the determination of possible positions for facial features, with higher risks of error.

   For eliminating the neck the system takes into account a range of possible ratios among the long and short axis of a face blob. On this range, the search is refined for the current subject. First, it is considered that most people present a narrower row in skin blob at neck level. Thus, starting from the ellipse center, the blob widest row is searched. Finally, the

narrowest blob row, that should be upper to the widest row, is located. A new ellipse is approximated to the cropped blob, see Figure 2.
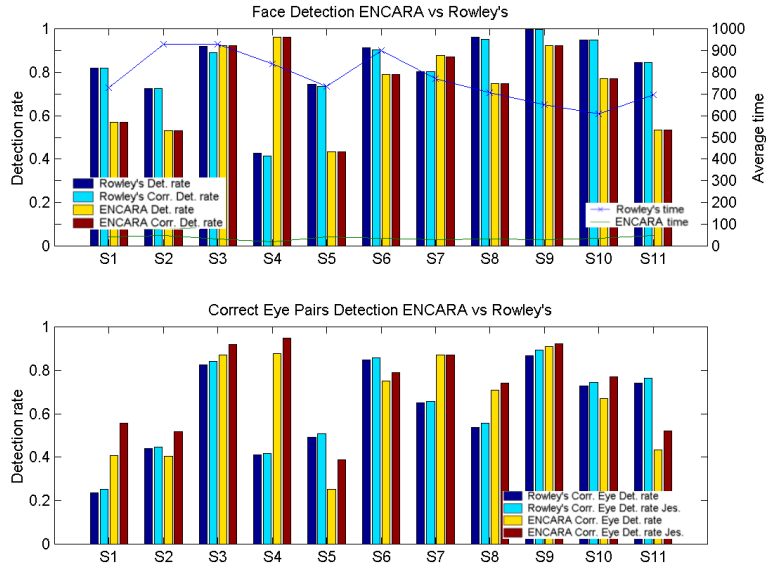

Figure 3. Results summary comparing ENCARA with Rowley's technique.

(e) *Eyes Detection:* At this point, the candidate has been rotated and cropped. As faces present geometric relations for features positions, the system searches each eye as a gray minimum in specific areas that have a coherent position for a frontal face. The search for each eye is bounded also by integrating integral projections. Each hemiface integral projection defines the upper y boundary of the search area, trying to avoid eyebrows. The minimal intereyes distance accepted in current version is 30 pixels.

(f) *Too Close Eyes Test:* If eyes detected using gray minima are too close in relation to ellipse dimensions, the closest one to ellipse center is refused. The search area $x's$ is modified avoiding the subarea where it was detected previously, searching the eye again.

(g) *Geometric tests:* Some tests are applied to gray level detected eyes:

   i. *Horizontal test:* Resulting eye candidates in transformed image should lie almost on a horizontal line if the ellipse orientation was correctly estimated. Using a threshold adapted to ellipse dimensions, candidate eyes that are too far from an horizontal line are refused.

   ii. *Intereye distance test:* Eyes should be at a certain distance coherent with ellipse dimension.

   iii. *Lateral test:* Face position is considered lateral if the distance from eyes to the closest border of the ellipse differs considerably.

4. *Normalization, M3:* A candidate set that verifies all the previous requirements is scaled and translated to fit a standard position and size (59×65). Finally, this normalized image is masked using an ellipse defined by means of normalized eye positions.
5. *Pattern Matching Confirmation, M4:* The appearance of the normalized image is tested in two steps.
   (a) *Eye appearance test:* A certain area (11×11) around both eyes in the normalized image is projected to a PCA eigenspace and reconstructed. The reconstruction error [3] provides a measure of its eye appearance, and could be used to identify uncorrect eye detections. The PCA eigenspace was built off-line using eyes marked manually from three different individuals.
   (b) *Face appearance test:* A final appearance test applied to the whole normalized image in order to reduce false positives makes use of a Haar based operator [6].

For candidate areas that have reached this point, the system determines that they are frontal faces. Then some actions are taken:

(a) *Between eyes location:* The middle position between the eyes is computed.
(b) *Mouth detection:* Once eyes have been detected, the mouth, i.e. a dark area, is searched down the eyes line according to intereyes distance. The located mouth position is accepted only if it fits the prototypical distance of the face from the eyes line.
(c) *Nose detection:* Between eyes and mouth, ENCARA searches for another dark area using gray values for detecting nostrils. From nostrils upwards the brightest point found is selected as the nose.

3 Experimental results

In order to carry out empirical studies of the system, different video streams were acquired and recorded using a standard webcam (320×240 pixels) at 15 Hz during 30 seconds. These sequences, labelled S1-S11, were acquired on different days without special illumination restrictions. The sequences, of 7 different subjects, cover different gender, face sizes and hair styles. Ground data were manually marked for each frame in all sequences for eyes and mouth center in any pose. This gives $11 \times 450 = 4950$ images manually marked. All the frames contain one individual but the pose is not restricted.

ENCARA performance is compared both with humans and a face detector system. On the one hand, the system is compared with manually marked or ground data providing a measure of facial features detection accuracy in terms of human precision. Whenever ENCARA detects a face, it provides eye positions. These returned positions are compared with the manually marked. A correct face detection is considered when for both eyes the distance to ground data or manually marked eyes is lower than a threshold that depends on the actual distance between the eyes, *ground_data_inter_eyes_distance*/8. This threshold

is more restrictive than the one presented in [5] where the threshold established is twice the one presented here. Therefore, even if a face is roughly detected but its eyes are not correctly localized, this face detection will be considered incorrect. On the other hand, an excellent and widely used automatic face detector for frontal and profile faces [7] has been applied on these images.

The results achieved with both face detectors are provided in Figure 3. Top graph shows for all the sequences, S1-S11, the detected faces and correct detected faces for both approaches. Bottom graph indicates the rates for correctly detected eye pairs according to two different criteria as described above. For each sequence, the first two bars are referred to Rowley-Kanade's approach, while the last two to ENCARA performance. The polylines plotted present the average processing time in milliseconds using the standard C *clock()* function in a PIII 1Ghz for both approaches.

According to this Figure, ENCARA performs for the worst case, S10, 16.5 times and in the best case, S4, 39.8 faster than Rowley-Kanade's technique. Calculating the average excluding the best and the worst times gives and average of 22 times faster than Rowley's technique. However, the face detection rate is worst for ENCARA except for S3, S4 and S7.

This performance for ENCARA is accompanied by a correct eye pairs location rate according to Jesorsky's criterium greater than 97.5% (except for S5 which is 89.7%). This rate is generally better than the one provided by Rowley's technique, this fact can be explained due to this technique does not provide eye detections for every face detected.

For most sequences Rowley's technique detects more faces, however that difference is reduced if the comparison is carried out with the number of faces detected with both eye positions. Previous Figure reflects a performance for ENCARA that detects an average of 84% of the faces detected using Rowley-Kanade's technique but 22 times faster using standard acquisition and processing hardware. ENCARA provides also the added value of detecting facial features for each detected face.

4   Conclusions and future work

The current system implementation presents promising results in desktop scenarios providing frontal face detection and facial features localization data valid to be used by face processing techniques. The main goal established in the requirements of the system was real time processing. The final experiments present a rate of 20-25 Hz for the image sizes used in the experiments, using a PIII 1Ghz. Its performance is much faster but still not so reliable in comparison with Rowley-Kanade's technique. This rates are reached thanks to the combination of different fast techniques such as skin color detection and tracking. The system has been designed to be updated and improved according to ideas and/or techniques that could be integrated.

The development of a real time robust face detector system is a hard task. There are many different aspects that can be considered to improve ENCARA

performance. Future work will necessarily pay attention to increasing color detection performance and adaptability. Current version depends on a first detection based on color which can be affected by illumination conditions and the existence of skin color in the background. In those situations the system avoids false detections thanks to the use of the appearance confirmation step.

Acknowledgments

References

1. R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings IEEE*, 83(5):705–740, 1995.
2. Gianluca Donato, Marian Stewart Bartlett, Joseph C. Hager, Paul Ekman, and Terrence J. Sejnowski. Classifying facial actions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(10), October 1999.
3. Erik Hjelmas and Ivar Farup. Experimental comparison of face/non-face classifiers. In *Procs. of the Third International Conference on Audio- and Video-Based Person Authentication. Lecture Notes in Computer Science 2091*, 2001.
4. Erik Hjelmas and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83(3), 2001.
5. Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz. Robust face detection using the hausdorff distance. *Lecture Notes in Computer Science. Procs. of the Third International Conference on Audio- and Video-Based Person Authentication*, 2091:90–95, 2001.
6. Stan Z. Li, Long Zhu, ZhenQiu Zhang, Andrew Blake, HongJiag Zhang, and Harry Shum. Statistical learning of multi-view face detection. In *European Conference Computer Vision*, 2002.
7. Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.
8. Karin Sobottka and Ioannis Pitas. A novel method for automatic face segmentation, face feature extraction and tracking. *Signal Processing: Image Communication*, 12(3), 1998.
9. Christopher Wren, Ali Azarrbayejani, Trevor Darrell, and Alex Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), July 1997.
10. Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.