ENCARA: REAL-TIME DETECTION OF FRONTAL FACES

M. Castrillón Santana, M. Hernández Tejera, J. Cabrera Gámez

Instituto Universitario Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería
Universidad de Las Palmas de Gran Canaria
35017 Gran Canaria - Spain
modesto@dis.ulpgc.es

ABSTRACT

This paper describes a real-time approach for face detection and selection of frontal views, for further processing. Typically, face detection papers provide results for a set of single images but the problem of face detection in video streams rarely is tackled. Instead of performing an exhaustive search for every video stream frame a set of opportunistic ideas applied in a cascade fashion and based on temporal and spatial coherence provide promising results in real-time.

1. INTRODUCTION

Current society is characterized by an incremental and notorious integration of computers in daily life, both in social and individual contexts. However, it happens that sometimes those machines that have been designed to help humans provoke rejection or stress among them. This is mainly due to the fact that Human Computer Interaction (HCI) is currently based on the use of certain devices or tools that are clearly unnatural for humans.

Human beings are sociable by nature and use their sensorial and motor capabilities to communicate with their environment. Humans communicate not only with words but with sounds and gestures. Gestures are really important in human interaction [1]. Thus, body communication, gestures, facial expression are used simultaneously with sounds produced by our throat.

Could a computer make use of this information? If HCI could be more similar to human to human communication, accessing these artificial devices could be wider, easier and they could improve its social acceptability as human assistants. This approach would make HCI non-intrusive, more natural, comfortable and not strange for humans [2].

In this paper it is described a module for real-time face detection designed for HCI applications. The system performance is compared with the original implementation of a well known single image face detector [3].

2. ENCARA FACE DETECTOR

Face detection systems described in the literature can be classified along different dimensions. One of them is based on the use of knowledge employed by these systems: implicitly or explicitly. The first focuses on learning a classifier from a training samples set, providing robust detection for restricted scales and orientations at low rate. These techniques perform with brute force without attending to some evidences or stimuli that could launch the face processing modules, similar to the way some authors consider that the human system works [4]. On the other hand, the second group exploits the explicit knowledge of structural and appearance face characteristics that could be provided from human experience, offering fast processing for restricted scenarios.

ENCARA merges both orientations in order to make use opportunistically of their advantages and conditioned by the need of getting a real-time system with standard general purpose hardware. ENCARA selects candidates using explicit knowledge for later applying a fast implicit knowledge based approach.

Classification is the nuclear process in face detection. There are multiple possible solutions, that roughly speaking can be sorted in two groups: Individual and Multiple classifiers. The complex nature of the face detection problem is easily addressed by means of an approach based on multiple classifiers. The proposed architecture for combination of classifiers follows [5] and is sketched in Figure 1. However, there is a main difference in relation to that work where the classifiers are based only on rectangle features [5], in this model the different nature of the classifiers used is assumed and promoted.

Initially, evidence about the presence of a face in the image is obtained and the face hypothesis is launched for areas of high evidence. A first classifier module confirms/rejects the initial hypothesis in the most salient area. If it is not confirmed, the initial hypothesis is immediately rejected and the classifier chain is broken, directing the system towards other areas in current image or to the detection of new ev-

idences. On the other side, if the hypothesis is confirmed, the following module in the cascade is launched in the same way. This process, for an initial hypothesis consecutively confirmed by all modules, is finished when the last module confirms also the face hypothesis. In this case, the combined classifier output is a positive face detection.

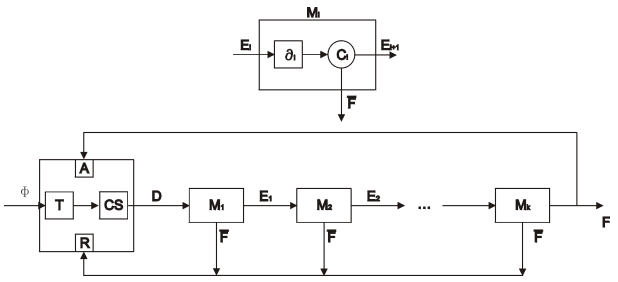

Fig. 1. T means tracking and CS Candidate Selection, D are data, M_i is the i-th module, C_i the i-th classifier, E_i the i-th evidence, A accept, R Reject, $F/\bar{F}$ face/nonface, ∂_i the i-th evidence computation and Φ the video stream.

The number of cascade stages and the complexity of each stage must be sufficient to achieve similar detection performance while minimizing computation. So, given a false positive rate, f_i, and the detection rate, d_i, for classifier module *i*, the false positive rate, FP, and the detection rate D of the cascade are respectively:

$$FP = \prod_{i=1}^{K} f_i \qquad D = \prod_{i=1}^{K} d_i \tag{1}$$

where K is the number of classifiers. These expressions show that cascade combination is capable to obtain good classification rates and very low false positive rates if the detection rate of individual classifiers is good, close to 1, but the false positive rate of them is not so good, not close to 0. For example, for $K = 10$ and a false positive rate of individual classifiers of 0.3, the resulting false positive rate is reduced to 6×10^{-6}.

This classifier combination scheme can be considered as a kind of pattern rejection or rejecter, in the sense given by [6], and can be interpreted in an analogy with fluid filtering in a filtering cascade. In this case, each filtering stage rejects a fraction of impurity. The more stages with a rejection rate, the more pure fluid is obtained at the output.

At this point a question raises: how to select the individual classifier modules? Certainly, there are different options. In this document, an opportunistic criteria is employed to extract cues and to use, in a convenient fashion, explicit and implicit knowledge to restrict the solutions to a solutions space fraction which can comply with real-time restrictions and have a flexible framework to test different solutions, adding modules or deleting another ones, allowing each module in the cascade to be also a combined classifier.

ENCARA is briefly described (detailed in [7]) in terms of the following main modules organized as a cascade of hypothesis confirmation/rejection schema :

M0.- Tracking: If there is a recent detection, the next frame is analyzed first searching for facial elements detected in the previous frame: eyes and mouth corners. If the tracked positions are similar to the one in the previous frame and the appearance test is passed, ENCARA considers that a face has been detected.

M1.- Face Candidate Selection: The current implementation makes use of a skin color approach to select rectangular areas in the image which could contain a face. Once the normalized red and green 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 structuring element.

M2.- Facial Features Detection: Frontal faces would verify some restrictions for several salient facial features. In those candidates areas selected by *M1* module, the system removes heuristically elements that are not part of the face, e.g. neck, and fits an ellipse to recover the blob vertical position. Later, this module searches for a first frontal detection based on facial features and its restrictions: geometric interrelations and appearance. This approach would first search potential eyes in selected areas taking into consideration that for caucasian faces, the eyes are dark areas in the face. After the first detection of an user, the detection process will be adapted to user dimensions and appearance as a consequence of temporal coherence enforcement.

M3.- Normalization: In any case, the development of a general system capable of detecting faces at different scales must include a size normalization process in order to allow for a posterior face analysis and recognition reducing the problem dimensionality.

M4.- Pattern Matching Confirmation: A final confirmation step of the resulting normalized image is necessary to reduce the number of false positives. This step is based on an implicit knowledge technique.

For eye appearance, a certain area (11×11) around both eyes is projected to a Principal Components Analysis (PCA) eigenspace and reconstructed. The reconstruction error [8] provides a measure of its eye appearance, and could be used to identify incorrect eye

detections. If this test is passed, a final appearance test applied to the whole normalized image in order to reduce false positives makes use of a PCA representation that is classified using Support Vector Machines.

If the tests are passed, the mouth and nose are located in relation to eye pair position and their dark appearance in a face.

In any other case, when no frontal face is detected, the system computes if there was a recent face detection at least one facial feature was not lost according to tracking process, the possible face location is estimated with high likelihood.

3. DETECTION EXPERIMENTS

In order to carry out empirical evaluations of the system, different video streams were acquired and recorded using a standard webcam. These sequences, labelled S1-S11, were registered on different days without special illumination restrictions. Each sequence contains 450 frames of 320×240 pixels. Ground truth data were manually marked for each frame in all sequences, for eyes and mouth center in any pose.

The sequences were analyzed using a PIII 1Ghz. As shown in Figure 2, ENCARA performs for the worst case, S10, 16.5 times and in the best case, S4, 39.8 times 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-Kanade's technique.

To verify the validity of those detections, two different criteria are used: 1) A face is considered correctly detected if both eyes and the mouth are contained in the rectangle returned by the face detector, and 2) The eye pair is considered correctly detected if for both eyes the distance to manually marked eyes is lower than $1/8$ the actual distance between the eyes. This threshold is more restrictive than the one presented in [9] where it is assumed twice this value.

ENCARA correct eye pairs location rate is greater than 80% according to according to the restrictive threshold and greater than 97.5% (except for S5 which is 89.7%) according to Jesorsky's criterium. This rate is always better than the one provided by Rowley-Kanade's technique, see bottom Figure 2, this fact can be explained due to the fact that this technique does not provide eye detections for every face detected. However, the face detection rate is worst for ENCARA except for S3, S4 and S7. In the worst case ENCARA detects only 58% of those faces detected using Rowley-Kanade's technique, S5. However, the average excluding the best and the worst performances is 83.4%.

ENCARA detects an average of 84% of the faces detected using Rowley-Kanade's technique, but 22 times faster

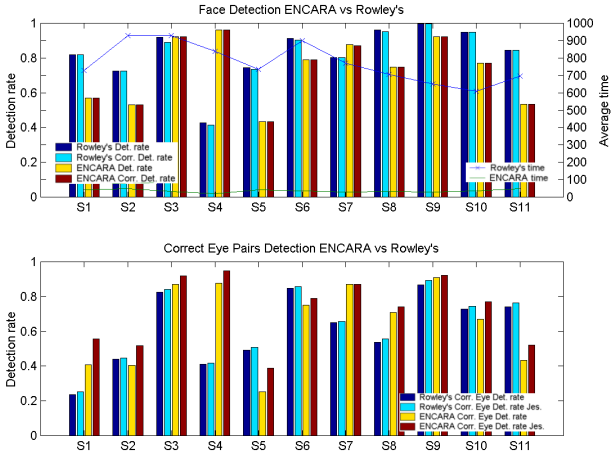

Fig. 2. Results summary comparing ENCARA with Rowley-Kanade's technique.

using standard acquisition and processing hardware. ENCARA provides also the added value of detecting facial features for each detected face.

Observing the system working live, sometimes it happens that the face changes and ENCARA is not able to track every facial element nor find again the face in that frame. In that situation, if there was a recent detection and at least one face feature was not lost, as mentioned above, ENCARA provides a likely location. The use of temporal coherence in this way allows ENCARA to track the face even when there is a sudden out of plane rotation, or the user blinks, etc. Figure 3 compares face detection rate results if these likely locations returned by ENCARA are considered as face detections, the image plots the face and the correct detection rates for Rowley-Kanade's and ENCARA.

As it is observed in that Figure, the ENCARA face detection rate increases significantly its performance keeping its frame rate. Performance becomes similar or better to that provided by Rowley-Kanade's algorithm. This rate keeps an excellent correct face detection rate which is always over 93.5%. It must be reminded that this rate considers as correct detections those faces in which both eyes and mouth are contained in the rectangle provided by ENCARA.

The ENCARA results can be summarized pointing out that ENCARA performs an average of 22 times faster than Rowley-Kanade's, detecting an average (excluding both best and worst case) of 95.2% of the faces detected by Rowley-Kanade's algorithm.

4. CONCLUSIONS AND FUTURE WORK

A solution to develop a real-time facial detector using standard hardware has been proposed. A cascade solution based

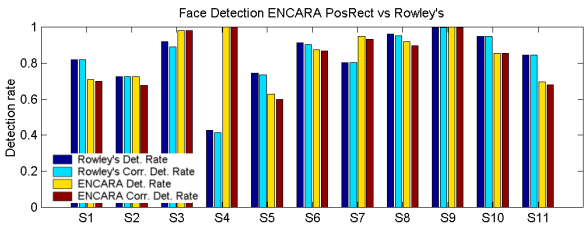

Fig. 3. Results summary comparing ENCARA considering Possible Rectangle as Face Detection with Rowley-Kanade's technique.

on weak and low cost classifiers of any nature is designed. The main features of the system are:

- The resulting system integrates and coordinates different techniques, heuristics and common sense ideas adapted from the literature, or conceived during the development.
- The system is based on a hypothesis verification/ rejection schema applied opportunistically in cascade, making use of spatial and temporal coherence.
- The system uses implicit and explicit knowledge.
- The current system implementation presents promising results in desktop scenarios providing frontal face detection and facial features localization data.
- The system has been designed in a modular fashion to be updated, modified and improved according to ideas and/or techniques that could be integrated.

ENCARA schema fulfills the real-time restriction with shorter detection time than Rowley-Kanade's technique, and its detection rates are in average slightly lower, though in some situations they are competitive. However, the current implementation is highly sensitive to lighting condition changes. More research must be done in order to improve the current candidate selection module by means of the integration of more cues in the cascade classifier. The integration of different weighted cues in parallel will improve robustness and reduce the dependence on an unique and sensitive classifier.

For HCI, the main aspect for face detection should not be the detection rate for individual images but to provide a good enough rate to allow HCI applications work properly. This has been demonstrated empirically with different sample applications developed for ENCARA [7].

Acknowledgments

We would like to thank Dr. H. Rowley for providing the binaries of his face detector.

The first author is supported by Consejería de Educación, Cultura y Deportes of the Comunidad Autónoma de Canarias, and Beleyma and Unelco through Fundación Canaria Universitaria de Las Palmas. This work was partially funded by research projects PI2000/042 of Gobierno de Canarias and UNI2002/16 of Universidad de Las Palmas de Gran Canaria.

5. REFERENCES

[1] David McNeill, *Hand and Mind: What Gestures Reveal about Thought*, University of Chicago Press, 1992.

[2] Alex Pentland, "Looking at people: Sensing for ubiquitous and wearable computing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, January 2000.

[3] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade, "Neural network-based face detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[4] Andrew W. Young, *Face and Mind*, Oxford Cognitive Science Series. Oxford University Press, 1998.

[5] Paul Viola and Michael J. Jones, "Robust real-time object detection," Technical Report Series CRL 2001/01, Cambridge Research Laboratory, February 2001.

[6] S. Baker and S.K. Nayar, "Pattern rejection," in *Proceedings of the 1996 IEEE Conference on Computer Vision and Pattern Recognition*, June 1996, pp. 544–549.

[7] Modesto Castrillón Santana, *On Real-Time Face Detection in Video Streams. An Opportunistic Approach.*, Ph.D. thesis, Universidad de Las Palmas de Gran Canaria, March 2003.

[8] 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.

[9] 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*, vol. 2091, pp. 90–95, 2001.