ARTÍCULO

Towards Real-Time Multiresolution Face/Head Detection[*]

M. Castrillón-Santana, H. Kruppa,[**] C. Guerra-Artal, M. Hernández-Tejera

Universidad de Las Palmas de Gran Canaria
Instituto Universitario de Sistemas Inteligentes
y Aplicaciones Numéricas en Ingeniería
Edificio Central del Parque Científico-Tecnológico
Campus Universitario de Tafira
35017 Las Palmas - España
mcastrillon@iusiani.ulpgc.es

Abstract

Reliable and real-time face detection is a basic ability for any Vision Based Interface. This paper combines and exploits the benefits of two different face detectors specialized each one in a specific context. The resulting system improves their respective individual performances by means of their cooperation, the integration of temporal coherence, persistence and explicit knowledge about the human face, achieving a robust and close to real-time multiresolution face detector.

Keywords: Face Detection, Real-Time, Human Computer Interaction.

1. Introduction

Human beings are sociable by nature and use their sensorial and motor capabilities to communicate. It is obvious that we communicate not only with words but also with sounds and gestures. If Human Computer Interaction (HCI) could be more similar to human to human communication, HCI would be non-intrusive, comfortable and not strange for humans [16]. Therefore, people detection is a basic ability to be included in any Vision Based Interface [21] in order to perceive the user in an HCI context.

Different approaches have been developed in the past for people detection attending to different elements of the human body: the face [7, 26], the head [1, 2], the entire body [24] or just the legs [15], as well as the human skin [10]. However, in this context it is obvious that the human face is a main information channel during the communication process [14].

[*] Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2003/06, UNI2004/10 and UNI2004/25, Canary Islands Autonomous Government PI2003/160 and PI2003/165 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

[**] Institute of Computational Science Haldeneggsteig 4 / Weinbergstrasse ETH-Zentrum, IFW C 29.1 CH-8092 Zürich, Switzerland, hkruppa@inf.ethz.ch

**Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. Vol. 9 No 27 (2005), pp. 63-72.
ISSN: 1137-3601. ©AEPIA (http://www.aepia.org)**

Any system devoted to the facial analysis must first detect the face to analyze. Face detection is a revisited topic in the literature with recent successful results [12, 18, 23]. However, these detectors focus on the problem using approaches which are valid for restricted face dimensions and, with the exception of the first reference, to a reduced head pose range.

In this paper, we design a real-time vision system which goes beyond traditional still image face detectors, adding to a state of the art object centered face detector two new elements in order to get a better, more robust, more flexible and real-time multiresolution face detector. These two new cues are: 1) the temporal coherence, and 2) the advantages evidenced by the local context in head detection for low resolution and difficult head poses [11]. These abilities extend the application of other face detection systems, building a system which is able to manage robustly not only typical desktop interactive applications but also surveillance situations and the transition between both contexts, i.e. face and head detection.

2. Face Detection

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 which are interesting for social interaction such as gender, age, expression and more. That ability allows us to socialize adequately with other people based on the information extracted visually from their face.

Face detection can be defined as *to determine any face -if any- in the image returning the location and extent of each* [7, 26]. In this paper, we have considered the information used to model faces to classify the different face detection techniques into two main families:

- Pattern based or Implicit: These approaches work searching exhaustively a previously learned pattern at every position and different scales of the whole input image.
- Knowledge based or Explicit: These approaches increase processing speed by taking into account face knowledge explicitly, exploiting and combining cues such as color, motion, face and facial features geometry and appearance.

Recent implicit face detectors [18, 23] have reduced dramatically the processing latency at high levels of accuracy. Particularly the Viola-Jones' object detector framework [22], has been recently made available integrated in OpenCV (Open Computer Vision Library) [8]. This framework, designed for rapid object detection, is based on the concept of a boosted cascade classifier [22] but extends the original feature set and provides different boosting variants for learning [13]. The resulting detection rate, D, and the false positive rate, F, of the cascade is given by the combination of each single stage classifier rates, i.e. d_i and f_i respectively:

$$D = \prod_{i=1}^{K} d_i \qquad\qquad F = \prod_{i=1}^{K} f_i \qquad (1)$$

Under this approach, given a 20 stage detector designed for refusing at each stage 50 % of the non-object patterns (false positive rate) while falsely eliminating only 0.1 % of the object patterns (target detection rate), its expected overall detection rate is $0{,}999^{20} \approx 0{,}98$ with a false positive rate of $0{,}5^{20} \approx 0{,}9 * 10^{-6}$. Therefore, the detector designer chooses the desired number of stages, the target false positive and detection rates per stage, achieving a trade-off between accuracy and speed for the resulting classifier.

3. Our Face Detection Approach

According to Torralba [20] object-centered approaches dominate the research in computational vision based face detection. Indeed, most face detection systems have been designed for a resolution range in which the face and its features are clearly distinguishable. This focus restricts the face detectors applicability due to the fact that those systems are easily fooled in situations with poor conditions in terms of pose, resolution, illumination, or occlusion. In those situations, humans make use of the context that plays a major role in order to achieve greater robustness.

The approach presented here, schematically described in Figure 1, combines and integrates, in a single real-time system, the benefits of two face detectors specialized each one for a different resolution range and, exclusively if a video stream is processed, the integration of temporal coherence by means of tracking.

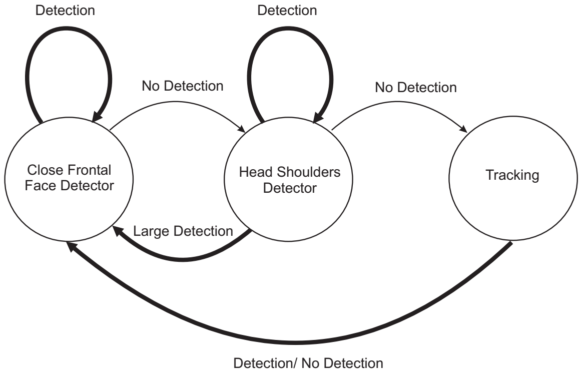

Figure 1. Combined face detector. A thick line represents that the next frame is processed while a narrow line indicates that the same frame is processed using another technique. Therefore a detection using the object centered detector will launch in the next frame the object centered detector centered in a Window of Attention defined by the previous frame detection. A detection provided by the local context based detector will launch in the next frame only the local context detector in the window of attention, unless the detection were big enough to contain a face detectable by the object centered detector.

Thick transitions in Figure 1 corresponds to situations where a new frame is acquired. Narrow transitions indicates that another technique is employed (i.e. no face has been found). For a given frame the system tries: 1) first the **object centered detector** which can additionally provide facial features, 2) if no face was located, a **local context based detector** is then employed, and 3) finally, if the system is searching a recently detected face but it could not be found by any of the previous techniques a **tracker** is used with the last face pattern available.

The system is designed trying to avoid the execution of both detectors for each frame, reducing time consumption, giving priority to the object-centered detector, i. e. the one which provides more facial details. Figure 1 shows that whenever one of the detectors detects, it takes the lead in the next frames until it fails. Common sense says that if there is a face, perhaps the shoulders are not present, but if the head and shoulders are detected, a head (thus perhaps a face) must be there. Thus, if the local context detector detects a pattern big enough to contain a detectable face, the object-centered detector is first applied in the next frame.

The approach assumes that a detected face is characterized by $f = \langle pos, size, red, green, leye_{pos}, leye_{pattern}, reye_{pos}, reye_{pattern}, face_{pattern} \rangle$. If there is a detection then this face features vector is created, and those features are used in the next frame to direct the search using the different integrated techniques mentioned in the previous paragraph. A detection is associated to a previous detection if it fits with the previously stored model.

The face detection information extracted from the previous frame is certainly useful to speed up the process, e.g., it is used to define a *Window of Attention* where the previous detection will likely be, or if the face size is big enough to be worth the application of the object centered detector. In any case, this information is valuable to reduce the time consumption. Obviously, if there were no recent detection, there would not be any active face model, and therefore the object-centered and local context detectors would be applied sequentially to the whole image.

The resulting system is able to manage in real-time complex scenes in which the human face experiences large scale, pose and appearance transformations. Each specialized detector is described in more detail below.

3.1. Object-centered detector

The object centered face detector is related to both previously mentioned families of face detectors, see Section 2, as it makes use of both implicit and explicit knowledge to obtain the best of each one. The implicit knowledge is integrated using the Viola-Jones framework [22] by means of the 25 stage face detector distributed with OpenCV.

On the other side, the explicit knowledge is added combining multiple simple classifiers of limited computational cost applied opportunistically in a cascade approach analyzing different areas of interest. A first detection given by the implicit face detector allows the system to have an estimation for the next frame of different parameters of the individual: his particular color model, his last position, etc. In the next frame, the process launches an initial face hypothesis on selected image areas, trying to confirm/reject that initial frontal face hypothesis. In the first case, the confirmation, the module results are passed to the following module. In the second, the rejection, the candidate area is rejected. Those techniques are combined and coordinated with temporal information extracted from a video stream to improve the performance, being able to provide a result faster than the original implicit face detector.

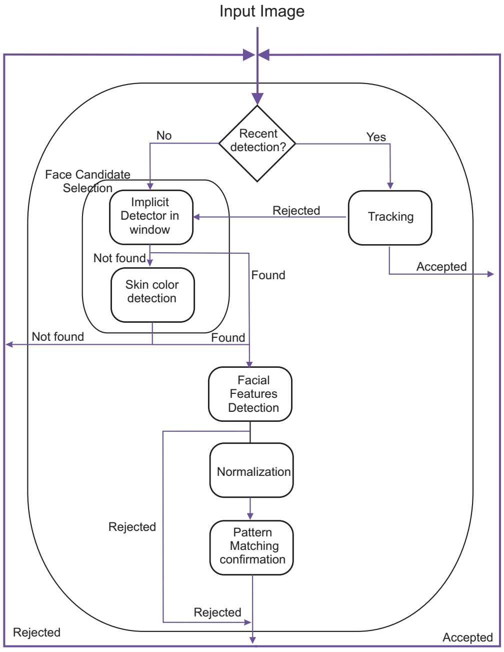

Figure 2. Object-centered face detector modules.

The resulting Frontal Face Detector developed is briefly described in terms of the main modules presented in Figure 2:

1. *Tracking, M0*: After a face is detected, if its facial elements were also located, in the next frame the system tries first to track the eyes, instead of detecting the whole face again. The search area for each facial element is related in size and position to the previous detection. Several 24×24 patterns are searched in the next frame using an approach based on the Sum of Squared Differences (SSD), similarly to [5]:

$$D(u,v) = \sum_{Area} |I(u+i, v+j) - P(i,j)| \quad (2)$$

   This approach decides autonomously whether the pattern must be updated or not, and if the pattern must be considered lost.

2. *Face Candidate Selection, M1:* This module combines two different focuses.

   a) *Implicit Knowledge Based Face Detection:* This step is similar to the face detector described in [23]. However there is a major change in order to speed up the process. By default the detector searches in the whole image, but if there was a recent face detection, the search area considered is three times the last detected size. A positive face detection will allow also to model the particular skin color of the individual. This skin color model is used to locate the face blob.

   b) *Skin Color Based Detection:* The skin color extracted from the face previously detected is used to locate the face if previous cues failed. The normalized red and green image [25] of the current frame is calculated to get those blobs that fit the current skin color model. The system does not make use of a general skin color model but a particular skin color model obtained after a detection. This is done to reduce the illumination dependence of a general skin color model description.

3. *Facial Features Detection, M2:* The detector searches facial features in the candidate skin blobs provided by *M1*. Major blobs are fitted to a general ellipse, refusing some of them by means of their dimensions. The orientation of the biggest one is used to rotate the source image forcing the eyes to lie on an horizontal line. Later, the system searches each eye as a gray minimum in specific areas coherent for a frontal face. Those positions achieved must be coherent with ellipse dimensions to be accepted.

4. *Normalization, M3:* A candidate eye pair set that verifies all the previous requirements is scaled and translated to fit a standard position and size.

5. *Pattern Matching Confirmation, M4:* Finally the appearance of the normalized image is tested in two steps:

   a) *Eye appearance test:* A certain area (11×11) around both eyes is projected to a PCA eigenspace and reconstructed. The reconstruction error [6] provides a measure of its eye appearance, and could be used to identify uncorrect eye detections.

b) *Face appearance test:* A final appearance test applied to the whole normalized image. The image is projected to a PCA space where a classifier based on Support Vector Machines confirms its facial appearance.

For a candidate area that reaches this point, the system determines that it contains a frontal face.

3.2. Local context detector

It is surprising to see how easily current state of the art object centered face detectors fail in situations in which humans have no problem in detecting faces reliably. The face detector described in the previous section, which fits the standard object-centered approach, fails for example in frames 129, 171, 230, 246, 278, 392, 430, and 564 of the Office sequence, see Figure 3. Such cases have been systematically studied in psychophysical experiments by Sinha and Torralba [19]. Their experiments indicate that humans make use of the *local context*, understood as a local area surrounding the face, as the level of detail decreases. At that resolution, the standard and predominant object-centered approach is not able to manage properly the problem.

The Viola-Jones' general object detection framework [22] has been used to build a head and shoulders detector. This approach combines increasingly more complex classifiers in a cascade, and has already been used recently for detecting frontal faces [23], pedestrians [24], eyes [3, 4] and faces at low resolution and difficult poses in [11]. Specifically, the head and shoulders face detector employed in [11] reported promising results applied to FGNET database providing a hint of head position, i.e. face. Here, we make use of the head and shoulders detector which allows the system to estimate the head location whenever the head and shoulders are present.

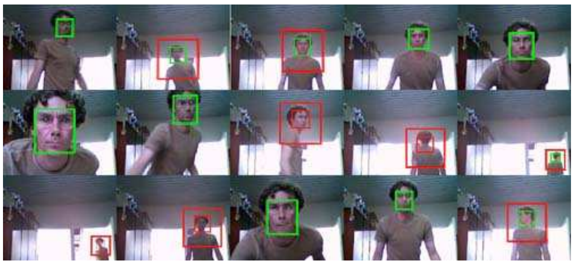

Figure 3. Frames 76, 129, 171, 180, 186, 206, 221, 230, 246, 278, 392, 430, 466, 502 and 564 with detections corresponding to the Office sequence, see footnote 2.

4. Experiments

Both detectors report high performance in their specific context. Here, we first describe some results achieved for each specialized detector in their context, and finally results are provided for situations where any single solution does not provide reliable performance. In all the experiments, test data sets are disjoint from the data used for training.

	Rowley		Viola		OCFD	
	TD	FD	TD	FD	TD	FD
Faces	88.8%	2.2%	90.1%	8.2%	92.9%	8%
Eyes	77.8%	1%	0.0%	-	64.3%	3.7%
Proc. time	422.4 msecs.		117.5 msecs.		21 msecs.	

Table 1. Results for face and eye detection processing 26360 **images. The correct detection ratios (TD) are given for the detections over the whole sequence, and the false detection ratios (FD) consider the total number of detections.**

4.1. Object centered detector

In Table 1, our object centered face detector (OCFD) is compared with the Viola-Jones' [22] and Rowley's [17] detectors. These results have been achieved processing 70 sequences which contain a total number of 26360 faces. In order to check the detectors performance, previously the sequences have been manually annotated, therefore the face containers are available for the whole set of images, and the eye locations are available for a subset of 4059 images.

In order to establish whether a detection is correct, two different criteria have been defined: 1) A face is considered correctly detected, if the detected face overlaps at least 80 % of the annotated area, and the area difference is not doubled. 2) The eyes of a face detected are considered correctly detected if for both eyes the distance to manually marked eyes is lower than a predefined threshold that depends on the actual distance between the eyes, *ground_data_inter_eyes_distance*/4 similarly to previously published papers [9].

On one side, the OCFD detection rate has a lower boundary which is given by the Viola-Jones' detector. We must also point out that OCFD provides the added value of facial features detection with slightly lower rates to those provided by the Rowley's approach but 20 times faster. On the other side, OCFD provides a processing cost similar to Viola-Jones' when there is no face, but the processing cost is reduced whenever faces are present in the video stream, as happens with the video streams used in these experiments. Therefore thanks to the use of temporal coherence in the next frames, the OCFD detector performed 5 times faster than the Viola-Jones' detector for these sequences. In the experiments the system spends and average of aprox. 20 msecs. to process an image. Our OCFD proves the advantages that temporal information contained in video streams provide to the problem of real-time face detection in video streams.

4.2. Local context detector

Several experiments have been carried out on FG-NET video conference data (PETS 2003) using the local context detector. Every 100th frame from sequences A, B, and D (cameras 1 and 2) is used in the following experiments[1]. This results in a total of 502 frames containing 1160 faces, 160 of which are profiles (about 14 %). The sequences show a conference room across either side of a desk with people occasionally entering and leaving the room.

In the experiments, the local context detector based on the Viola-Jones' framework has been compared with an object based face detector [13] which is now part of OpenCV [8]. The left plot in Figure 4 shows the face detection performance on the FGNET data set in terms of the ROC curve. Both the performance of the object-centered and the local context detector are shown. The percentage of retrieved faces is given on the vertical axis and the number of false positives per frame is shown on the horizontal axis. Points on the curve are sampled by varying the number of stages in the cascade.

The frame resolution is decreased from the left plot to the right plot. Each frame is downsampled from the original resolution (720 × 576 which approximately corresponds to PAL resolution) to 360 × 288 ("half-PAL") and to 180 × 144 pixels ("quarter-PAL"), respectively. Accordingly, the available face resolution decreases from 48×64 to 24 × 32 and to 12 × 16 pixels approximately. Figure 5 shows an example frame where each row corresponds to a different resolution (all frames have then been resized for visualization purposes).

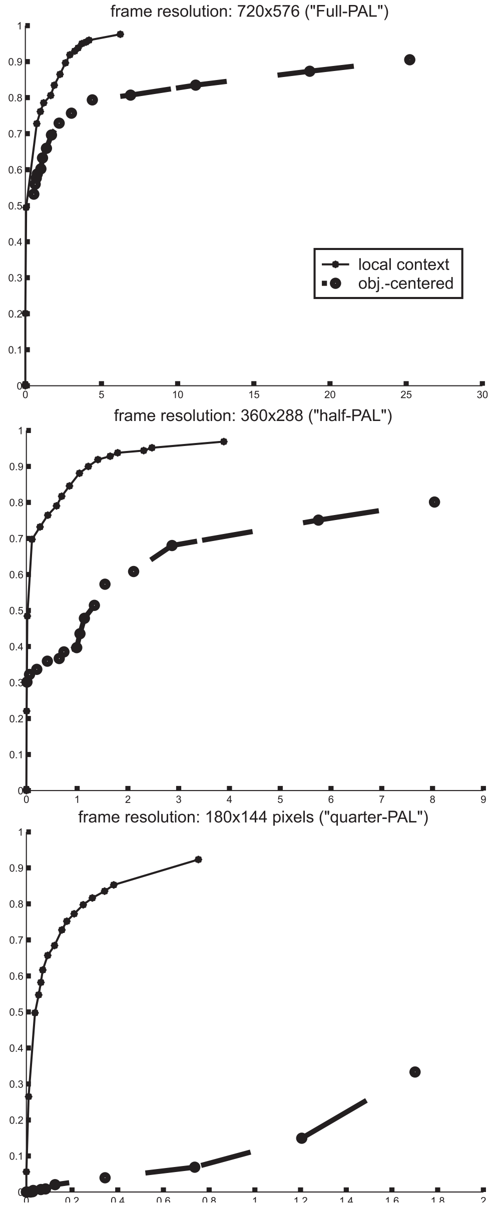

Figure 4. Detection accuracy on the FGNET data set. Each plot shows the percentage of detected faces (vertical) vs. the number of false positives per frame (horizontal). The ROC

[1] A similar subset was used in [3].

curves show the performance of the object-centered detector and the local context detector.

At the original resolution, the local context detector dominates already because it is more robust to face pose changes, see Figure 6. At 5 false alarms the object-centered detector retrieves 80 % of the faces and the curve flattens from there on. Contrastingly, the local context detector yields 95 % of the faces at the same level of false alarms.

At lower frame resolutions (middle plot and right side plot) facial details deteriorate and the object-centered approach becomes unreliable. The local context on the other hand is not affected. At half-PAL resolution the object-centered at 5 false alarms drops 15 % compared to the full resolution, affected by the decrease in available facial detail. Contrastingly, the local context detector's performance remains stable at 95 % given the same number of false alarms. This effect becomes even stronger at quarter-PAL resolution where the object centered approach detects only 10 % at 1 false alarm per frame while the local context detector succeeds for more than 90 % of the faces.

Overall the local context detector provides improvements in detection rates by 15 %, 25 % and 80 % at corresponding levels of false alarms. Additionally, since the local context detector can operate robustly at very low frame resolutions it actually runs 15 times faster than the traditional object-centered approach at the same level of accuracy.

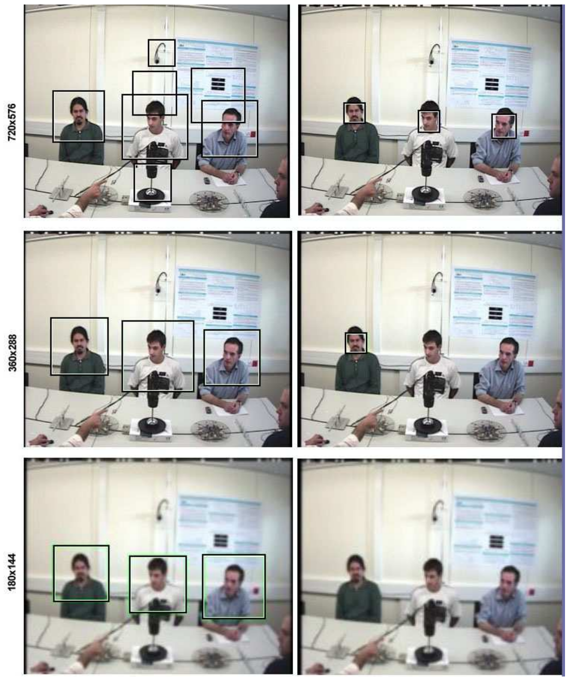

Figure 5. FGNET sequence: Each row of this figure corresponds to a different frame resolution (frame resolution decreases from the top to the bottom row). The images have been rescaled to the same size only for illustration purposes. The left column shows detections based on local context, the right column is from the object-centered approach. This example illustrates that as facial details degrade, the object-centered approach misses actual faces.

4.3. The Combined Face/Head Detector

Here, we tackle the combination of both detectors, as described in Section 3, in order to develop a robust, reliable and fast face/head detector for any context. Following the schema described in Figure 1 a prototype has been built combining both specialized detectors. In order to speed up the process, when there is no detection for a consecutive number of frames, instead of applying both detectors to the whole image for every frame, they alternate.

We have observed the different kind of detections provided by these detectors. The object-centered detects frontal faces and under some circumstances the added value of its main elements, i.e. eyes and mouth. The local context detector detects head and shoulders and therefore implicitly the head position.

In this prototype, the tracker is only used for local context detections, because they have a reduced size more suited to this technique. However, the integration of tracking introduces a risk, due to the fact that a wrong pattern can be adopted for tracking as the false detection rate is low but it exists. We have avoided most false detections analyzing their persistence in time. It was observed that false detections are typically isolated in time. In our experiments we have especially analyzed the local context detections likelihood by means of coherent detection persistence. Consecutive low resolution detections, with similar position and size, are considered as an evidence of a coherent head and shoulders detection, and therefore can be used as a pattern to track in the next frame.

Table 2 presents the performance of each single classifier and the combined detector for the Office sequence[2]. This sequence crosses the border between close and far contexts. In this table, a face is considered correctly detected if annotated eye positions are inside the rectangle estimated by the face detector and the ratio between its width and the intereyes distance is well proportioned. In the table, the low false detection rate achieved by both detectors is observed. But additionally, the impressive benefits that the local context detector provides for head detection at low resolution and with 3D pose changes are clearly visible. The resulting solution is able to detect not only standard human frontal faces but also frontal or backward silhouettes containing head and shoulders as for example those in frames 129, 171, 230-392 and 506, see Figure 3.

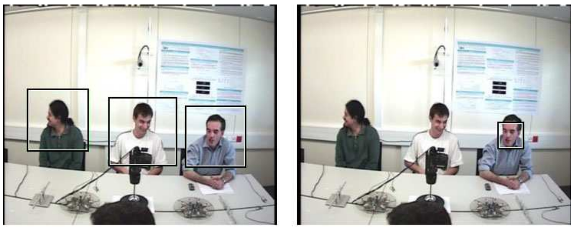

Figure 6. FGNET sequence, face pose changes: The left plot shows detections of the local context of faces while the right plot shows the output of the object-centered detector. The latter misses two faces because it is restricted to frontal view detection.

The combined detector improves clearly the results provided by an state of the art object centered face detector, for unrestricted sequences where there are large face changes in terms of scale, appearance and pose, just due to that detector is not able to manage those changes. It also improves the local context detector results being able to detect faces when the local context is not visible. The average processing time for the Office sequence reports 18 fps, but it must be pointed out that there is no face in 670 frames. When this situation does not happen, as for example for the desktop sequences, the detector performs at 24fps.

Detector	Average proc. time	Det. rate	False det. rate
Object Centered	23 msecs.	30.5%	0.0%
Local Context	82 msecs.	66%	1.4%
Combined	54 msecs.	81.8%	0.3%

Table 2. Results for the Office sequence, see Figure 3 and footnote 2. Average processing time considering 320×240 frames acquired with a standard webcam and a PIV 2.2Ghz.

5. Conclusions and Future Work

We have developed a robust face/head detector that integrates two specialized face detectors and temporal coherence at different levels for video streams processing. The resulting live demo system manages different facial resolutions and the transitions between them in close to real-time, 18fps, with fixed or mobile cameras because detection is not based on motion.

Current implementation manages just a single face (indeed the biggest in the image), our next project is to improve the system in terms of managing concurrently a larger number of simultaneous faces which would make the system a valid and very powerful detector for a wide range of applications. In that sense both detectors will be improved mainly to detect also lateral poses. The local context detector has been trained with frontal samples ignoring lateral ones which could be added to the training set to get a more general detector. We also plan to integrate other cues as for example color to detect close faces which present a non frontal view, exploiting the persistence of detection concept for the object-centered detector too. Therefore additional features will be integrated to improve performance.

Referencias

[1] Stan Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 232–237, June 1998.

[2] Marco La Cascia and Stan Sclaroff. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(4):322–336, April 2000.

[2] This sequence contains 1802 frames of an individual presenting a large range of scale, appearance and poses. His head appears completely in 1139 of them, available at ftp://mozart.dis.ulpgc.es/pub/misc/Office.avi and Office_results.avi.

[3] David Cristinacce and Tim Cootes. A comparison of two real-time face detection systems. In *Fourth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-ICVS), Graz, Austria*, pages 1–8. IEEE, April 2003.

[4] I. R. Fasel, B. Fortenberry, and J. R. Movellan. Gboost: A generative framework for boosting with applications to real time eye coding. *Computer Vision and Image Understanding*, pages 182–210, April 2005.

[5] Cayetano Guerra Artal. *Contribuciones al seguimiento visual precategórico*. PhD thesis, Universidad de Las Palmas de Gran Canaria, Octubre 2002.

[6] 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*, pages 65–70, 2001.

[7] Erik Hjelmas and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83(3):236–274, 2001.

[8] Intel. Intel Open Source Computer Vision Library, b4.0. www.intel.com/research/mrl/research/opencv, August 2004.

[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*, 2091:90–95, 2001.

[10] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. Technical Report Series CRL 98/11, Cambridge Research Laboratory, December 1998.

[11] Hannes Kruppa, Modesto Castrillón Santana, and Bernt Schiele. Fast and robust face finding via local context. In *Joint IEEE Internacional Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 157–164, October 2003.

[12] 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*, pages 67–81, 2002.

[13] Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM'03*, pages 297–304, Magdeburg, Germany, September 2003.

[14] Christine L. Lisetti and Diane J. Schiano. Automatic facial expression interpretation: Where human-computer interaction, artificial intelligence and cognitive science intersect. *Pragmatics and Cognition (Special Issue on Facial Information Processing: A Multidisciplinary Perspective*, 8(1):185–235, 2000.

[15] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of the International Conference on Computer Vision*, pages 555–562, 1998.

[16] Alex Pentland. Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 107–119, January 2000.

[17] 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.

[18] Henry Schneiderman and Takeo Kanade. A statistical method for 3d object detection applied to faces and cars. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1746–1759, 2000.

[19] Pawan Sinha and Antonio Torralba. Detecting faces in impoverished images. AI memo 2001-028, CBCL memo 208, Massachussets Institute of Technology, 2001.

[20] Antonio Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2):169–191, 2003.

[21] M. Turk. Computer vision in the interface. *Communications of the ACM*, 47(1):61–67, January 2004.

[22] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[23] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):151–173, May 2004.

[24] Paul Viola, Michael J. Jones, and Daniel Snow. Detecting pedestrians using patterns of motion and appearance. In *Proc. of the International Conference on Computer Vision*, volume 2, pages 734–741, October 2003.

[25] 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):780–785, July 1997.

[26] 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.