Combining face and facial feature detectors for face detection performance improvement

M. Castrillón-Santana, D. Hernández-Sosa, and J. Lorenzo-Navarro*

SIANI
Universidad de Las Palmas de Gran Canaria, Spain
mcastrillon@iusiani.ulpgc.es

Abstract. In this paper, we experimentally study the combination of face and facial feature detectors to improve face detection performance. The face detection problem, as suggeted by recent face detection challenges, is still not solved. Face detectors traditionally fail in large-scale problems and/or when the face is occluded or different head rotations are present. The combination of face and facial feature detectors is evaluated with a public database. The obtained results evidence an improvement in the positive detection rate while reducing the false detection rate. Additionally, we prove that the integration of facial feature detectors provides useful information for pose estimation and face alignment.

1 Introduction

Automatic facial analysis has become practical thanks to the robustness of recent reliable face detectors. Once the face is located, different applications, such as head pose, face aligment or gaze estimation, require its facial feature localization. Face alignment for instance, is indeed a necessary step before any further facial analysis. Though some authors do not observe any improvement using the face alignment in certain problems [11], its necessity is suggested by others [13], particularly if the classification is related to shape information [5, 9].

However, face detectors are still unreliable in different hard scenarios where the pose and illumination are not controlled [8, 12], or when a large-scale problem is tackled [4]. Therefore, face and facial element detection keeps being a common topic in the Computer Vision literature. Among the wide and recent literature on face detection, the Viola-Jones face detector [15] has received lots of attention within the community. The authors designed indeed a general object detection framework that requires a previous training stage. This stage is accomplished using a large set of roughly aligned samples of the object to detect (positive samples) and of images not containing the target (negative samples). To create a new classifier, positive and negative samples gathering, data annotation, data preparation and training phases must be accomplished.

* Work partially funded by the Spanish Ministry of Science and Innovation funds (TIN 2008-06068), and the Computer Science Department at ULPGC.

Thanks to the OpenCV library [7], the training tools are available to a large community of researchers. As a result of their work, different classifiers specialized in face processing have been made publicly available [7] and compared [2]. The available implementation [10], and its successful results applied to face detection, have made the Viola-Jones framework based detectors to be frequently used as a baseline.

Face detectors are typically designed for the frontal face configuration. Therefore, their performance is sensitive to both in-plane and out-of plane rotations. In this paper, we study experimentally available face detectors, with the aim at proving that by combining both global face detection and local feature detection, the overall face detection performance can be substantially improved. In this sense, the reader must take into consideration the existence of works focused on the reduction of false positives [1].

Different datasets have been designed to analyze face detection performance. Most of them contain single faces in a reduced set of poses. As a remarkable exception, we would like to mention the CMU image database [14]. This dataset contains a collection of heterogeneous images, feature that from our point of view, allows for a better evaluation of the classifier performance. More recently, initiatives such as FIW [6] have introduced new challenging situations to test the performance of the face related detectors with much larger and heterogeneous imagery. Fortunately, the availability of annotation data referred to the face and facial feature location, such as those provided by FDDB [8], supports the comparison of face detectors. FDDB is used in our experiments because it includes a larger number of annotated faces in unrestricted situations.

Next section summarizes the Viola-Jones detection framework. Later, the experimental setup and results achieved are presented, finishing the paper with the conclusions.

2 Viola-Jones based face and facial feature detection

Among the recent face detection approaches, we have selected the Viola-Jones object detection framework [15] for our experiments considering the public availability of face and facial feature detectors. However, we would like to make the reader evident that the combination approaches described below could be applied to any detection framework. Our emphasis is on the combination of cues not on the particular detector employed.

As mentioned above, the Viola-Jones framework requires a previous training stage, that makes use of a large set of positive roughly aligned samples of the object to detect, and images not containing the target. Both sets provide information about the target appearance space and its boundaries.

This training stage creates a boosted cascade of linear combinations of weak classifiers achieving a performance similar to a strong classifier, but reducing the processing cost (high for a strong classifier). The main idea behind the architecture is to waste less processing time in areas that are easy to classify. In fact, the reduction of processing cost allows the integration of object detectors based on this framework, in real-time applications.

To get a useful detector, these weak classifiers must be able to detect the target most of the time, and reject at least around half of the non target images. Once configured, a cascade of K weak classifiers under this architecture offers a target detection rate, D, and a false detection rate, F, that are given by the combination of each single stage weak classifier rates:

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

For example, assuming a cascade composed of 20 weak classifiers with a true detection rate, d_i, of at least 99% and a negative detection rate, f_i, not greater than 50%, its expected overall detection rate is 0.9920 with a false positive rate of $0.520 \approx 0.9 * 10^{-6}$. A reduction in the number of stages increases both rates, i.e. is less restrictive, and reduces computational cost.

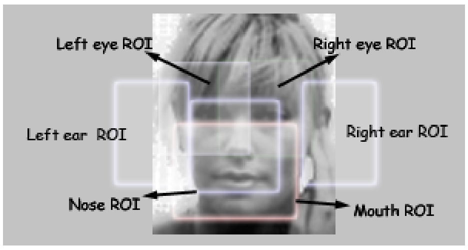


Fig. 1. Illustration of the ROIs areas defined in a face detection container to search the corresponding facial features. Left and right labels are related to the image.

3 Experimental results

3.1 Test dataset

The facial annotation data provided with the FDDB dataset has been made available in terms of ellipses [8]. The authors suggest the use of a score based on the Jaccard index, to determine when a detection must be considered a true or positive detection. The match degree between a detection det_i and an annotation $anot_j$ is given by:

$$S\left(det_i, anot_j\right) = \frac{area\left(det_i\right) \bigcap area\left(anot_j\right)}{area\left(det_i\right) \bigcup area\left(anot_j\right)}$$

A large match means that both intersection and union overlap in a high degree. In our experiments we have considered that det_i is a positive detection when $S(det_i, anot_j) > 0.5$ [8]. The number of annotated faces contained in the face dataset, i.e. FDDB, is 5171.

3.2 Detection results

Among the different face and head detectors included in the OpenCV release, we have chosen for our experiments FA2, labeled $haarcascade\_frontalface\_alt2$ in the OpenCV distribution. This detector presents a good detection rate and speed, achieving a larger Area Under the Curve (AUC) in [2].

For facial feature detection, we have selected those exhibiting the best performance in [2, 3] for eye, mouth, nose and ear detection. All of them are currently included in the OpenCV release. Our hypothesis is that we can get better performance introducing different heuristics in the face search. In this subsection we will compare different detection strategies:

- **F**: Face detection is performed using the FA2 classifier.
- **FC**: Face detection is performed using the FA2 classifier. Later, facial feature detection is applied within their respective expected Region of Interest (ROI), related to the detected face, see Figure 1. The fail in detecting at least four facial elements is used as a filter to remove likely false face detections. As a result, both positive and negative detection rates will be reduced. The different ROIs used, considering that sx and sy are respectively the width and height of the face container, are:
  - Left and right eyes: The left upper corner of their respective ROIs are $(0, 0)$ and $(sx \times 0.4, 0)$, their dimensions $(sx \times 0.6, sy * 0.6)$.
  - Nose: Left upper corner of the ROI are $(sx \times 0.2; sy \times 0.25)$ and its dimensions $(sx \times 0.6, sy \times 0.6)$.
  - Mouth: Left upper corner of the ROI are $(sx \times 0.1; sy \times 0.4)$ and its dimensions $(sx \times 0.8, sy \times 0.6)$.
  - Left and right ear: The left upper corner of their respective ROIs are $(-sx/3; sy \times 0.2)$ and $(sx/2; sy \times 0.2)$, and their dimensions $(sx/3 + sx/2, sy \times 0.6)$.
- **FC2**: This approach is similar to **FC**, but the face container is scaled up before searching the facial features. Ideally, the positive detection rate will be increased because the facial elements appear in more detail than in **FC**.
- **FFs**: No face detector is used, but facial feature detection is employed instead. The coocurrence of at least three coherently located detections gives support to a face presence. The basic rules applied to determine the coherence of two facial features detected are summarized as:
  - The mouth must be below any other facial feature, but not too far away.
  - The nose must be below both eyes, but above the mouth.
  - The centroid of the left eye must be to the left of any other facial feature and above nose and mouth.
  - The centroid of the right eye must be to the right of any other facial feature and above nose and mouth.
  - Ears must be on each side.
  - The separation distance must be coherent with the element size.
- **FFFs**: Combines **F** and **FFs** detection results, building a single set of detected faces. The objective is to be able to detect faces with hidden elements or slightly rotated, which are not easily removed using the **F** approach.

- **FCFFs**: Combines **FC** and **FFFs** to reduce the false positive rate.
- **FC2FFs**: Combines **FC2** and **FFFs** to both reduce the false positive rate, and increase the positive detection rate.
- **XXR**: All the previous approaches are also applied not only to the input image but to two slightly rotated images, ±15 degrees, to cope with more variations in the face pose.

The detection results obtained are presented in Table 1. We must point out that the FDDB annotation available is not completely exhaustive, i.e. no every face has been annotated. To avoid any artifact in the false detection rate, we have not considered as false detections those that are indeed non annotated faces. Their influence is particularly remarkable for facial features detection based approaches, as it is confirmed that if automatically considered false detections are revised by hand the false detection rate decreases drastically.

Table 1. True and false positive detection rates, respectivelly TPR and FPR, achieved for each approach.

Approach	TPR	FPR	Approach	TPR	FPR
F	0.7117	0.0470	FR	0.7401	0.1151
FC	0.6384	0.0015	FCR	0.6693	0.0052
FC2	0.6892	0.0041	FC2R	0.7169	0.0126
FFs	0.5007	0.0019	FFsR	0.5691	0.0052
FFFs	0.7248	0.0462	FFFsR	**0.7561**	0.1079
FCFFs	0.6511	0.0033	FCFFsR	0.6817	0.0044
FC2FFs	0.7008	0.0054	FC2FFsR	**0.7289**	0.0112

The baseline given by the selected face detector reports a positive detection rate around 71%, and a negative detection rate around 5%. The criterion used to accept a detection is identical to the one used in [8], i.e. if the ratio of the intersection of a detected region with an annotated face region is greater than 0.5, a score of 1 is assigned to the detected region, and 0 otherwise.

Observe that when a face is validated only if at least two inner facial features are located (**FC** and **FC2**), both rates decrease, but the false positives decrease drastically. This suggests the importance of a simple heuristic on the face detection performance. Indeed, the **FC2** approach achieves a similar positive detection rate, while reducing more than ten times the false detection rate.

Table 1 includes also results achieved making use of an approach that detects faces based on face features detectors, **FFs**. We have accepted a valid face only if at least three inner features are detected. The results reported are clearly worse in terms of positive detection. Indeed only half of the faces are located, however the false detection rate is remarkable low. The rest of the table indicates the results achieved if both focuses are combined with the aim at improving the overall rate. The behavior is similar if faces are confirmed by means of its inner features or not. Some detection samples with and without the integration of facial features detection in the process are depicted in Figure 2.

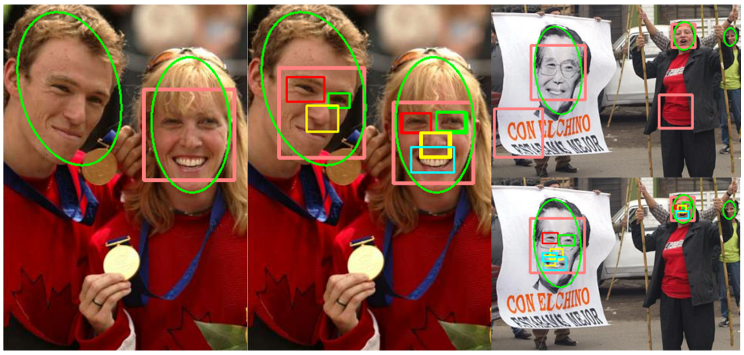
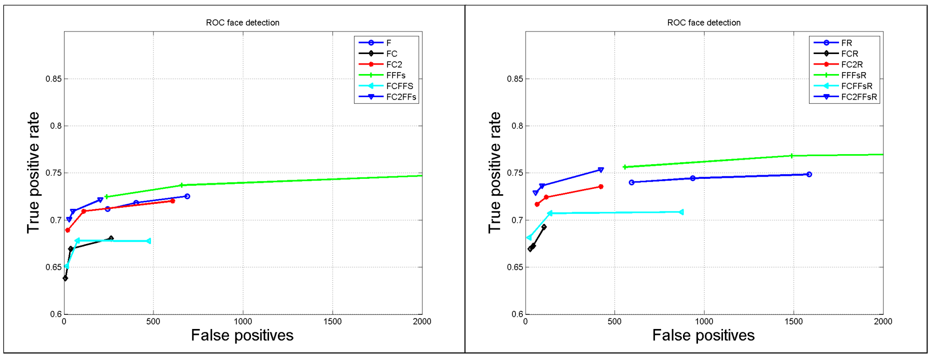


Fig. 2. Face detection examples based on **F** (left) and **FC2FFs** (right). They illustrate the benefits of the integration of facial features detectors, serving to remove false detections (upper row) and undetected faces located by its facial features (bottom row).

If the search is applied not only in the input image but also in slightly rotated images the results achieved improve the detection rate. The best performances are achieved when face and facial features detection are combined. Compared to the **F** approach the improvement is evident, however we must remind the reader that there is an additional cost due to the fact that multiple detectors are employed, but it may be considered almost irrelevant thanks to the current multicore architectures.

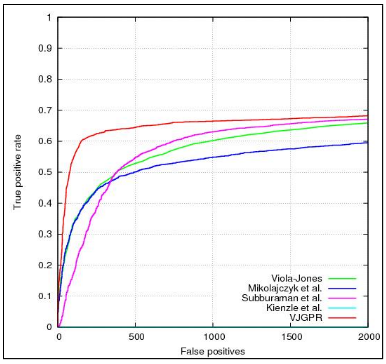


(a) (b)

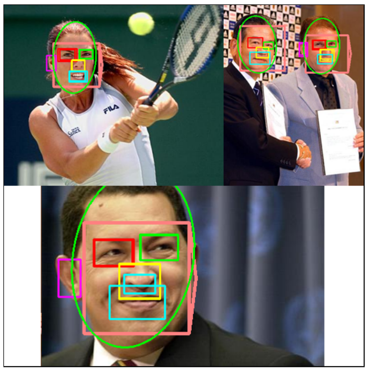
Fig. 3. Face detection results on the FDDB for the proposed approaches: (a) not including the search in rotated images, and (b) including the search in rotated images.

To illustrate better the benefits, we have computed, for each approach excepting those involving only face features, the receiver operating characteristic (ROC) curve applying first the original face classifier, and two variants reducing its number of stages. Theoretically, this action must increase both correct, D, and false, F, detection rates. The results are presented in Figure 3a. Similar graphs are shown in Figure 3b for the approaches that search the pattern in the original and rotated images. They present a slight improvement, but increasing three times the processing cost. The comparison with the results achieved in [8] for FDDB, see Figure 4a, evidences the better behavior exhibited by the approaches that combine face and facial features detectors. The improvement for 500 false positives is close to 10 percentage points.

An additional advantage of the integration of facial feature detectors in the process is that they provide useful information for other tasks. For example, the ear detection gives rough information of the head pose, as depicted in purple Figure 4b. Both ear detectors are designed for profile head poses. Thus, if just an ear is detected it could mean that the head is partially or totally rotated, we decide that according to the number of eyes detected for that particular face.



(a)



(b)

Fig. 4. (a) Face detection results on the FDDB for different approaches (extracted from [8]). (b) Examples of pose estimation based on the facial features detected.

4 Conclusion

We have presented an experimental study on the FDDB of a set of face and facial features detectors based on the Viola-Jones framework. The study considers different face, head and facial feature detectors. We focus on the benefits that their combination based on common sense heuristics would bring in the face detection problem. In this way, we have reduced the number of false face detections by locating inner facial features within face containers, increased the number of detected faces by means of combining with facial feature based detection, and with searching in slightly rotated images. The results achieved suggest a better performance than other state of the art detectors. Facial feature detection provides not only the possibility of further face confirmation and help with slightly rotated faces, but also information about the face pose.

References

1. C. Atanasoaei, C. McCool, and S. Marcel. A principled approach to remove false alarms by modelling the context of a face detector. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–11, 2010.
2. Modesto Castrillón, Oscar Déniz, Daniel Hernández, and Javier Lorenzo. A comparison of face and facial feature detectors based on the violajones general object detection framework. *Machine Vision and Applications*, 22(3):481–494, 2011.
3. Modesto Castrillón, Daniel Hernández, and Javier Lorenzo. An study on ear detection and its applications to face detection. In *Conferencia de la Asociación Española para la Inteligencia Artificial*, 2011.
4. Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. Large-scale privacy protection in google street view. In *In Proceedings of ICCV*, page 2373 2380, 2009.
5. B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *International Journal of Computer Vision Research*, 74(2), August 2007.
6. Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, October 2007.
7. Intel. Open Source Computer Vision Library, v2.3. http://opencv.willowgarage.com/wiki/, July 2011. (last visited May 2012).
8. Vidit Jain and Erik Learned-Miller. FDDB: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, 2010.
9. Andreas Lanitis, Chris Taylor, and Timothy F.Cootes. Automatic interpretation and coding of face image using flexible models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), July 1997.
10. Rainer Lienhart, Alexander Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM'03, 25th Pattern Recognition Symposium*, pages 297–304, Magdeburg, Germany, September 2003.
11. Erno Mäkinen and Roope Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):541–547, March 2008.
12. J. Parris, M. Wilber, B. Helfin, H. Rara, A. El-barkouky, A. Farag, J. Movellanand M. Castrillón-Santana, J. Lorenzo-Navarro, M.N. Teli, S. Marcel, C. Atanasoaei, and T. Boult. Face and eye detection on hard datasets. In *IEEE IAPR International Joint Conference on Biometrics (IJCB)*, 2011.
13. A. Rabie, C. Lang, M. Hanheide, M. Castrillón, and G. Sagerer. Automatic initialization for facial analysis in interactive robotics. In *6th International Conference on Computer Vision Systems, Vision for Cognitive Systems*, pages 517–526, 2008.
14. 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.
15. Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):151–173, May 2004.