# 1

# The Performance of the Haar Cascade Classifiers Applied to the Face and Eyes Detection

Adam Schmidt[1] and Andrzej Kasinski[2]

[1] Institute of Control and Information Engineering str. Piotrowo 3a 60-965 Poznan, Poland Schmidt.Adam@gmail.com
[2] Institute of Control and Information Engineering str. Piotrowo 3a 60-965 Poznan, Poland Andrzej.Kasinski@put.poznan.pl

**Summary.** Recently we have presented the hierarchical face and eye detection system based on Haar Cascade Classifiers. In this paper we focus on the optimization of detectors training. Moreover, we compare the performance of Lienhart's face detectors [1] and Castrillon-Santana's eyes detectors [2] with those which have been trained by us.

## 1.1 Introduction

The serial way of information processing in such face recognition systems results in cumulating errors of consecutive processing steps. Campadelli et al. [3] discuss the influence of the eyes localization error on face recognition (FR) ratio of some methods. He also concludes that some published FR results do not clearly report the fact of manual initialization performed, which results in better than average recognition rates reported by particular authors.

Anyway, in order to develop an automatic face recognition system, one has to design an efficient face and eye localization method.

In [4] we have presented hierarchical face and eyes detection system based on Viola's Haar Cascade Classifiers (HCC) [5] augmented by some knowledge-based rules.

Our main goal was to train the reliable face and eyes detecting HCCs and to use them in the above mentioned hierarchical system. Therefore, two different strategies for creating the training sets were tested. Moreover, the influence of weak classifiers complexity and the desired detection ratio of particular stage classifier on the overall performance of the detection system were assessed.

As neighboring positive responses of HCC are merged into a single detection, the influence of constraining minimum number of neighbors on the systems efficiency was investigated.

In this paper we present results of our tests. Firstly, we describe a new database of 10000 face images with manually marked face and eyes positions for reference. Secondly, we compare the results of face detection by using original Lienhart's detectors [1] with those which were trained on our image base. Then we compare the results of eye detection using Castrilon-Santana's HCC [2] with our custom eye HCCs. The trials have been performed for eyes detection on both manually selected and automatically detected ROIs of the faces.

## 1.2 The environment

In order to get statistically significant results, our classification method performance evaluation was conducted on a set of face images consisting of almost 10000 images of 100 people . The images were acquired in partially controlled illumination conditions, over uniform background, and stored as 2048x1536 pixels JPEG files. Each person's pictures were taken in the following sequences, while:

- turning their head from the right to the left,
- nodding their head from the raised to the lowered position,
- turning their raised head from the right to the left,
- turning their lowered head from the right to the left,
- moving their head without any constraint on the face pose.

The main purpose of creating such an image base was to provide an extensive and credible data for the systematic performance evaluation of the face detection, facial features extraction and face recognition algorithms. The number of gathered pictures guarantees sufficient inter- and between-class variability to obtain statistically reliable performance estimates of tested algorithms. In order to provide ground truth for the face detection and eyes extraction tasks the rectangle ROIs containing face and eyes were manually marked on each image in the base . The coordinates and dimensions of the rectangles circumcircling face and eyes were saved in OpenCV Storage files in the YAML format.

We implemented our system in the Visual C++ 6.0 by using the OpenCV [6] with Lienhart's implementation of the HCC. All detectors were trained using tools included in the OpenCV.

## 1.3 Detectors training

The training a HCC involves the application of large and diverse sets of positive and negative samples (images ). With our new image base it was easy to get faces and eyes examples. However, creating a set of "non-faces" is a tricky task, because what does it mean to represent every possible non-face object? As all images in our base were taken with the same background, we wanted to check whether the negative set built from the same images with hidden faces is sufficient enough to distinguish between faces and non-faces. Another negative training set was created by randomly gathering about 3500 diverse pictures not containing any faces. Negative example sets for eyes detector were created by hiding left or right eye on the avaible face image. The following face detectors have been trained:

- C1 - hidden faces negative set, 20x25 window, level min hit rate 0.995 and stump as a weak classifier
- C2 - hidden faces negative set, 20x25 window, level min hit rate 0.995 and 2-split CART as a weak classifier
- C3 - hidden faces negative set, 20x25 window, level min hit rate 0.995 and 4-split CART as a weak classifier
- C4 - rich negative set, 20x25 window, level min hit rate 0.990 and 4-split CART as a weak classifier
- C5 - rich negative set, 20x25 window, level min hit rate 0.995 and 4-split CART as a weak classifier
- C6 - rich negative set, 20x25 window, level min hit rate 0.999 and 4-split CART as a weak classifier

Only single parameter at a time was changed during the experiments . The variant giving best results was used in the subseqent tests. The C1, C2 and C3 detectors differ in the complexity of weak classifiers used. The C3 was trained with the hidden faces negative set, while the C4 was trained using the rich negative set. The C4, C5 and C6 HCCs vary in the required TP ratio of a single stage classifiers.

Their performance was compared to the following Lienhart's detectors already avaible in the OpenCV:

- L1 - stump based, 24x24 window, Discrete AdaBoost
- L2 - stump based, 20x20 window, Gentle AdaBoost
- L3 - 2-split CART based, 20x20 window, Gentle AdaBoost
- L4 - 2-split CART based, 20x20 window, Gentle AdaBoost, with a tree made of stage classifiers instead of a cascade

Independently, the following eyes detectors were trained and the results compared with Castrillon-Santana's detector results (S):

- E1 - level min hit rate 0.995, with 4-split CART as a weak classifier
- E2 - level min hit rate 0.999, with 4-split CART as a weak classifier

Each of our detectors was trained with the Gentle AdaBoost to the theoretical FP ratio of 10e-6 .

## 1.4 Results

### 1.4.1 Face detection results

We applied Lienhart's and our detectors to the whole image base. Minimal detection window's size was 400x500 pixels for our detectors and 400x400 for Lienhart's. As Lienhart's detectors were trained on square windows, their outputs aspect ratio was reduced to 4:5 for compatibility. If intersection area of both the detected and ground truth rectangles was greater than 80% of both rectangles' areas a TP was claimed, otherwise the case was considered to be a FP. If no face was found on a whole picture, it was declared as false negative (FN). Influence of a minimum number of the neighbors on the face detection ratio was tested. The average time of the image processing with each detector was measured.

**Table 1.1.** Average time of face detection on a PC with Intel Celeron 2800 MHz processor and 512 MB RAM

| Detector | Average detection time [ms] |
|---|---|
| C1 | 246.07 |
| C2 | 230.09 |
| C3 | 227.49 |
| C4 | 227.26 |
| C5 | 214.15 |
| C6 | 221.29 |
| L1 | 503.93 |
| L2 | 281.57 |
| L3 | 237.87 |
| L4 | 492.72 |

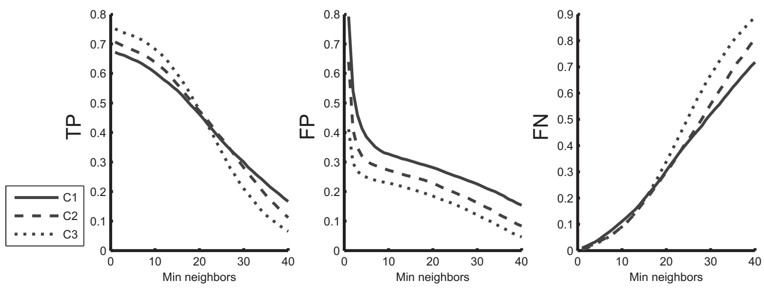


**Fig. 1.1.** The influence of weak classifier's complexity on face detection ratios

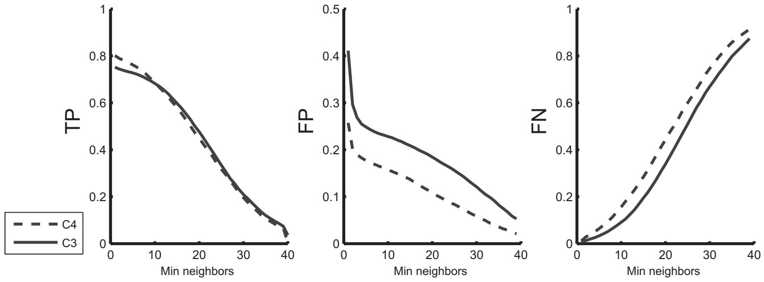


**Fig. 1.2.** The influence of the training set diversity on face detection ratios

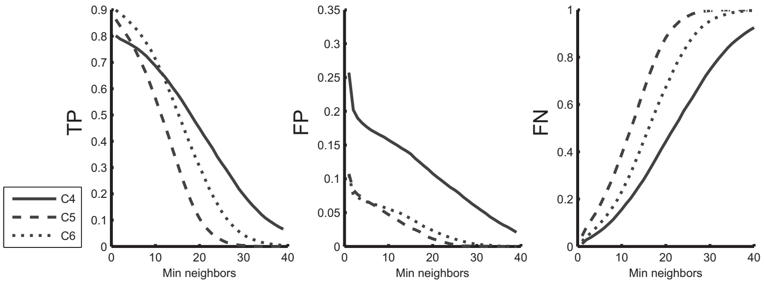


**Fig. 1.3.** The influence of the desired stage classifier's TP ratio on face detection ratios

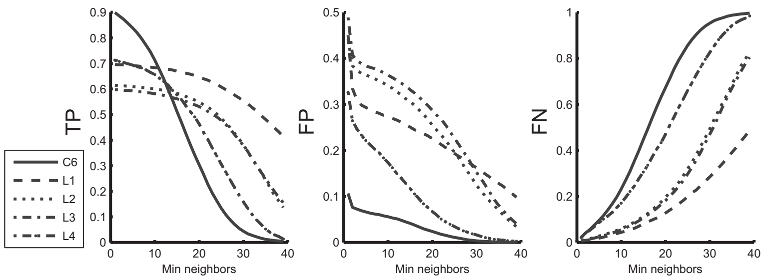


**Fig. 1.4.** The comparision of Lienhart's face detectors with our best HCC

### 1.4.2 Eyes detection results

We compared the results of the Castrillon-Santana's detectors with ours on the manually marked faces and the faces detected with C3, C6, L1 and L4 detectors. Error metric used here was the same as that of Campadelli[3]:

$$error = \frac{\max(\|C_l - C_{lGT}\|, \|C_r - C_{rGT}\|)}{\|C_{lGT} - C_{rGT}\|} \tag{1.1}$$

Where $C_l$ stands for the center of the left eye found, $C_r$ stands for the center of the right eye found, $C_{lGT}$ and $C_{rGT}$ for the centers of ground truth eyes.

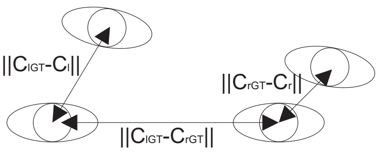


**Fig. 1.5.** Distances used in calculating the eye detection error

Detections with relative error below 0.1 were treated as TPs, the others were considered as FsP. Pictures without any positive eyes detection were counted as false negatives (FNs).

Subsequently, the influence of minimum neighbors constraints of both face and eyes detectors on the eyes detection rate was tested. The average processing time of eyes detectors in both regionalized (R) and non-regionalized (NR) search was measured.

**Table 1.2.** Average detection time for eyes on a PC with Intel Celeron 2800 MHz processor and 512 MB RAM

| Detector | Average detection time [ms] |
|---|---|
| E1 NR | 377.35 |
| E1 R | 105.83 |
| E2 NR | 337.59 |
| E2 R | 100.06 |
| S NR | 2171.08 |
| S R | 625.07 |

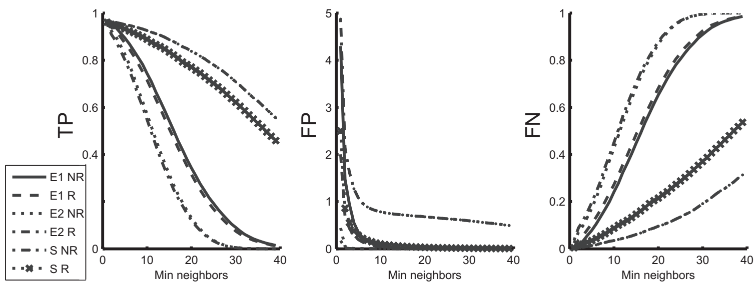


**Fig. 1.6.** Eyes detectors performance while face ROIs have been manually marked beforehand

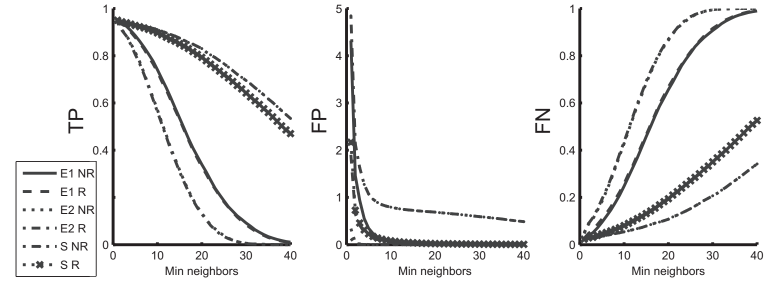


**Fig. 1.7.** Eyes detectors performance while face ROIs have been determined with the C6 face detector

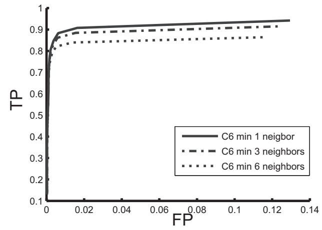


**Fig. 1.8.** TP ratio vs FP ratio of E2 R detector while face ROIs have been determined with different minimum neighbors constraint set

## 1.5 Conclusions

Our tests proved that using more complex weak classifiers pays off resulting in much more reliable detections. By increasing the number of splits in CARTs used as weak classifiers, higher TP ratio and lower FP ratio w.r.t the face detection were obtained. Many of FPs could be removed by restricting the minimum number of merged detections. However, it should be pointed out that only increasing it to the number of 5 produced positive results. Further increase of required neighbors resulted in the quick deterioration of the TP ratio without any significant change in the FP ratio.

Detectors trained on the plain background images turned out to be less efficient than those trained on rich negative set. This suggests, that in order to build a reliable classifier it is recommended to use as diversified negative set as possible.

By using 4-split CART as a weak classifier and setting the required TP ratio of each stage for 0.999 we were able to train the detector which outperformed all Lienhart's detectors both w.r.t. the detection ratio and computational efficiency. By using solely the face detector we were able to detect 90% of the faces, getting the FP ratio of 11%.

Our results confirmed the hypothesis that using the regionalized search results in a significant reduction of the FP ratio and the processing time. Castrillon-Santana's and our detectors reached comparable TP ratio, however our proposal turned out to have a several times lower FP ratio. It is worth to admit that processing time with our detectors was also 6 times shorter.

The eyes' detection ratios, while restricted only to the face ROIs, were comparable despite various face detectors involved. It's worth noting, that the best results were achieved with no restrictions on the minimum neighbors number of the face detector. This can be explained by the hierarchical nature of our system. Reducing the selectiveness at the first level (face detection) results in a propagation of dubious candidates. Thus more regions have to be evaluated at the second stage (eyes detection), which subsequently would discard face candidates with no eyes found. As the overall detection ratios of such a cascade are multiplicative, the TP ratio stays close to 100%, while the number of FP is reduced greatly. By using the combination of both our face and our regionalized eyes detector we were able to fully automatically detect eyes in 94% of images still keeping the FP ratio of 13%. By applying minimum neighbors constraint solely to the eyes detector the TP ratio of 88% was achieved with less than 1% FP.

## References

1. Lienhart R, Kuranov A, Pisarevsky V (2002) Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. Technical report, Microprocessor Research Lab, Intel Labs
2. Castrillón-Santana M, Lorenzo-Navarro J, Déniz-Suárez O and Falcón-Martel A (2005) Multiple face detection at different resolutions for perceptual user interfaces. In: 2nd ICPRIA. Estoril, Portugal
3. Campadelli P, Lanzarotti R, Lipori G (2006 (preprint)) Eye localization: a survey. In: The fundamentals of verbal and non verbal communication and the biometrical issues. NATO Science Series
4. Kasinski A, Schmidt A (2007) The Architecture of the Face and Eyes Detection System Based on Cascade Classifiers. Submitted to CORES 2007
5. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of CVPR 1:511–518
6. The Open Computer Vision Library. sourceforge.net/projects/opencvlibrary/