Scientific Bulletin of the „Petru Maior" University of Târgu Mureş
Vol. 7 (XXIV) no. 1, 2010, ISSN 1841-9267

COMBINING BOOSTED GLOBAL- AND PART-ASPECT FACE DETECTORS

Szidónia Lefkovits[1]
[1]"Petru Maior"University of Târgu Mureş, Romania
szidonia.lefkovits@science.upm.ro

ABSTRACT

The domain of object detection presents a wide range of interest due to its numerous application possibilities especially real time applications. All of them require high detection rate correlated with short processing time. One of the most efficient systems, working with visual information, were presented in the publication of Viola et al. [1], [2].
This detection system uses classifiers based on Haar-like separating features combined with the AdaBoost learning algorithm. The most important bottleneck of the system is the big number of false detections at high hit rate. In this paper we propose to overcome this disadvantage by using specialized parts classifiers.
This aim comes from the observation that the target object does not resemble the false detections at all.
The reason of this fact is the coding manner of Haar-like features which attend to handle image patches and neglect the edges and contours.
In order to obtain a more robust classifier, a global aspect method is combined with a part-based method, having the goal to improve the performance of the detector without significant increase of the detection time.

Keywords: part-based object detection, classifier, face and facial feature detection

1. Introduction

Object detection is one of the most widespread research domains in Computer Vision. The newest trends tend to detect several objects or classes of object in a single image.

In fact this complex problem has to be traced back to many binary classification problems. For each target object a separate specialized classifier has to be built up.

There are two main trends in the object detection: the global aspect methods and the part-based methods. The first ones have the advantage of quick detection, but can hardly handle occlusion, clutter or deformation. To the contrary the part-based methods overcome the drawbacks mentioned above by splitting the objects in parts and modeling, in addition, the connections between them. In such a way the object detection becomes more flexible, because the final decision depends on the joint probability of the components and connections.

Our idea is to combine the global aspect methods and the part-based methods in order to improve the detection rate and to decrease the number of false positives.

The significant number of false positives is obtained due to the global aspect based methods which can detect only fixed-size objects. Accordingly, exhaustive search is used over the image in order to handle the variety of scale and position of the object.

We propose, as a last step of the detection process different part-based classifiers. The role of these is the detailed analysis of hard examples based on different types of features. Our experiment uses Haar-like features and the AdaBoost learning algorithm as a global classifier and part-based classifiers consisting of first subband Wavelet coefficients and a kind of part-structure and probabilistic decision. The experiments were made on faces as target objects.

2. State of the art

There are lots of methods suggested for object detection. They can be distinguished by considering the feature set and the learning algorithm. Visual information can be extracted in a row form considering only image fragments. These features are simple and not very informative. So a large number of features are required in order to be sufficiently discriminative. In this case a complex classifier has to be elaborated, because of the high dimension of the feature space.

Schneiderman et al.'s [3] object detection system

works with many wavelet coefficients obtaining a complex decision rule, based on their probabilities.

Vidal-Naquet et al. [4] combine informative features with linear classification. His features are image fragments selected by maximizing several information criteria.

Leibe et al. [5] obtain a classifier based on a bag of words created from image patches. The object model thus consists of patches and their relative position to the object-center. Employing the similarity between patches, the decision is made by a statistical voting space.

Fergus et al. created a generative probabilistic model which expends the probabilities of aspect, position and size of several object parts in order to obtain the most likely configuration of the parts in the object. The learning phase identifies the correspondence between parts of the same object in different images.

Felzenszwalb et al.'s [7] face detection system is based on images containing labeled object parts. Out of these a pictorial structure model is built. The measure of the matching of the object with this structure is evaluated as a minimization of a cost function.

The most famous system for face detection was proposed by Viola et al.[1], [2]. Here the global aspect of faces is characterized by a set of Haar-like features and the discrimination of objects is based on a boosted set of classifiers.

This approach is further improved by Lienhart [8], [9] with the creation of a multi-stage general object classification system. The open source implementation [10] of this system is suitable for further applications in the domain.

In his paper Castrillón-Santana[11] analyzed the performance of all face and facial feature classifiers based on the above mentioned system. But the presented performance of the facial feature classifiers still needs some improvement.

The usage of these facial feature detectors has been studied by Wilson et al. [12] who propose to eliminate the false detections with a simple geometric consistency.

On the other hand Schulz [13], proposed for Viola et al.'s algorithm a final stage based on neural network for pedestrian detection.

3. System overview

Our approach has two major phases: in the first phase an exhaustive search is made for the target object considering it like a whole. In the second phase the detection is concentrating on the position of particular features. The classifiers are generated out of a training data set by a statistical learning algorithm AdaBoost. The AdaBoost algorithm was proposed by Freund and Shapire [14]. It constructs an ensemble of classifiers and uses a voting mechanism for the classification. The idea of boosting is to use the weak classifier to form a highly accurate prediction rule by calling the weak classifier repeatedly on different distributions over the training examples. The most important theoretical propriety of AdaBoost concerns in its ability to reduce the training error. The AdaBoost converts a set of weak classifiers into a strong learning algorithm, which can generate an arbitrarily low error rate. The weak classifiers (1) are built out of Haar rectangular features.

$$h(x,y) = sign\big(f_j(x,y) - \theta_j\big) \cdot p_j \qquad (1)$$

$f_j$represents the response of the Haar function, $\theta_j$is the best separating threshold of the Haar values over the training set and $p_j$is a parity value.

Significant is the very fast evaluation of them using Integral Images at different size and position with a scanning window.

The value of the Haar function is the measure of likelihood between a specified subregion of an image and the graphical representation having the same size of the rectangular Haar function, $RF_{Haar}(x_0,y_0,x,y)$

$$f_j(x_0,y_0) = \sum_{(x,y)\in I} I(x,y) \times RF_{Haar}(x_0,y_0,x,y) \qquad (2)$$

The result of the product (2) of an image region $I(x,y)$ with the rectangular Haar feature is the mean intensity of the underlying pixels. The obtained values code the patches efficiently. One source of the multitude of false positives can be explained by coding the patches with their mean intensity.

Another way is to consider only contours instead of patches.

In order to reduce the computation time, it is necessary to evaluate simple images very fast and only difficult decidable images have to be analyzed in detail. This aspect is solved efficiently by a cascade of classifiers. The cascade design process is driven by a set of detection performances [8]. If each stage classifier is taught for low performances ($f < 0.5$, false detection rate/stage and $d > 0.999$, hit rate/stage), then the whole cascade will have the same performances as a monolithic classifier, but much faster.

Each stage consists of a strong classifier built from weak classifiers h_t. The weight α_t of each classifier in the final classifier is determined by the AdaBoost algorithm.

$$H(I) = sign\left(\sum_t \alpha_t h_t - \frac{1}{2}\sum_t \alpha_t\right) \qquad (3)$$

Each stage is taught with the remaining images from previous stage. The stop condition of the learning process is given by the reached performance.

In order to satisfy the hit rate of the whole cascade classifier we need a very high hit rate for each strong classifier of it. Owing to the exhaustive search

over every scale and position in an image there are about 100,000 - 1 million windows to evaluate. This implies that the false detection rate of the cascade classifier has to be less than 10^{-6} To obtain this value for a $s = 20$ staged classifier we need for each stage a false detection rate of $f = 0.5$. The resulting false detection rate, $F = f^s$, is the product of false detection rate of the component stages. But simultaneously the hit rate is decreasing with each new added stage, $D = d^s$, that is why the stage hit rate has to be almost 1. In real systems this criterion can hardly be satisfied.

The overall training process involves two types of tradeoffs. In most cases, classifiers with more features will achieve higher detection rates and lower false positive rates. At the same time, classifiers with more features require more time to compute.

Fig. 1. *Compared image patches(-Equalized Wavelet/Gradient/Raw patches)*

Our proposal is to build, as a last stage, a specialized cascade classifier with the imposed high detection rate and to eliminate the false detections by training only on hard positive examples.

The last stage has to consist of many different part classifiers. A modality to build this is out of fully labeled training examples. We propose to use different types of classifiers based on raw patches, gradient images and wavelet images. The most important advantage of such a hierarchical classifier is the unnecessity of high performances. It can be stated that the performance is decreasing with the classifiable part dimension. Hence the high resolution of images is indispensable, so that all parts carry sufficient distinguishable information. The restriction for each specific part classifier related to its performance specifies only the high detection rate regardless the number false positives. This performance can be reached by modifying the classification threshold. The characteristic of the final classifier can be compensated by an adequate part-structure. This is represented by a collection of parts with connection between pairs. An instance of a part in an image is specified by a location X_i. We consider all possible connections between pats. Each connection is a deformation function measuring how well the location X_i agrees with the object model. The best fit of the configuration is computed from the joint probabilities of the detected positions (4) relative to the ideal location (see figure (2)).

$$P(x) = \mathrm{N}(x_{Le}|\mu_{Le}, \Sigma_{Le})\mathrm{N}(x_{Re}|\mu_{Re}, \Sigma_{Re}) \cdot \cdot \mathrm{N}(x_n|\mu_n, \Sigma_n)\mathrm{N}(x_m|\mu_m, \Sigma_m) \quad (4)$$

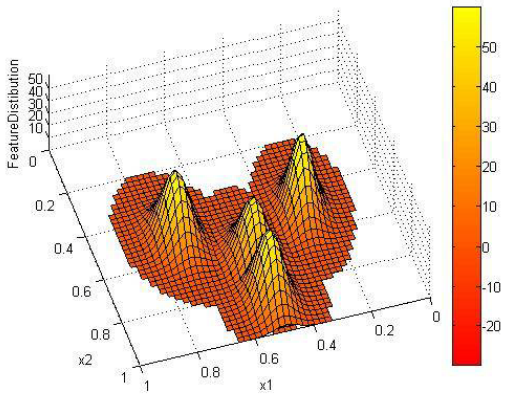


Fig. 2. *Measured probability distribution*

We define similar functions for different detection possibilities: three and two part detections also. The distribution of each part *le*-left eye, *re*-right eye, *m*-mouth, *n*-nose) position x is modeled by a normal distribution determining its own parameters (mean and covariance) experimentally from the training data set.

We define similar functions for different detection possibilities: three and two part detections also. The decision threshold for each different configuration has to be determined experimentally using the training data set and the obtained hard examples.

The model (see figure (3)) is built from a fully labeled training data set.

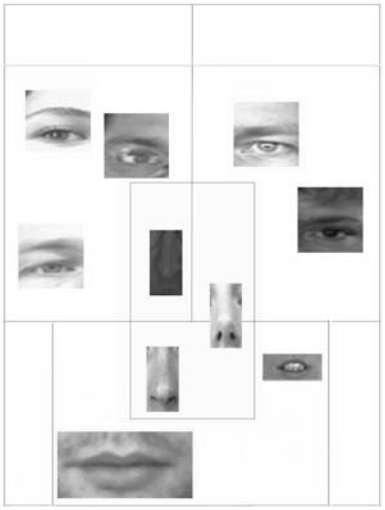


Fig. 3. *Part-structure*

3. Experimental results

In previous papers [15], [16] we discussed practical consideration in order to build a classifier. Our goal was to build a face classifier. The training images were selected from public, labeled face databases [17], [18] and completed with self-marked cropped studio images.

In order to solve the problem of false detections we proposed an algorithm which generates a high number of negative examples necessary at each stage of the training process.

To evaluate our classifier we used the test images offered by CMU [19]: it contains 105 images with 368 faces. The performance of this classifier can be evaluated from the ROC curve (see figure (4)). The detection rate obtained in early stages shows the properness of the face database used. This is due to the prepared pictures which do not contain sufficient various faces. They predominantly present young European people without beard or moustache, and very few of them wear glasses.

As we can observe from the ROC curves of our classifier (see figure (4)), with each new stage the false detection decreased considerably, but the detection rate decreased too. From the 9th stage the decrease of detection rate is significant. In addition, the number of weak classifiers/stage increases with each new stage. An enough high hit rate with this measurement is reached at stage 9.

At this stage our algorithm finds 2230 false images in the whole test set (see table (1)). Analyzing the detection results we can observe that the false images do not resemble the faces at all (see figure (5)).

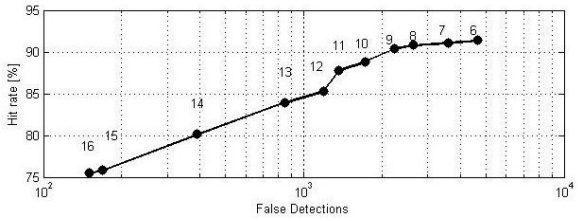


Fig. 4. *Stage by stage ROC curve*

Table1. *Measured performance of classifiers*

stage no.	6	7	8	9	10	11	12	13	14	15	16
Hit rate	91.4	91.1	90.8	90.4	88.85	87.8	85.32	83.96	8.16	75.8	75.5
No. false det.	4661	3588	2646	2230	1722	1363	1188	846	390	168	150
No. of cls./stage	214	283	387	460	557	672	765	869	1056	1242	1319

In our experiments two types of classifiers (raw patches and wavelet patches) are compared (see table (3)). For each facial feature: left eye, right eye, nose, mouth (see figure (1)) a separated classifier was built (see table (2)). The performance of these classifiers is indicated in table (3). Each of them evaluates only its specific region of interest (see figure (3)). This area was limited according to the face structure and moreover takes in to consideration the aspect of the resulted images in the previous stage. In order to build our part-structure in the last step the images are evaluated at a fixed aspect ratio (90×120).The parameters are learned from the wholly labeled training data set and the decision of the detection is made comparing the probability function (4) with an experimentally determined threshold. The threshold is modified according to the number of detected parts.

Table2. *Measured performance of part-classifiers*

Detections	Left Eye	Right Eye	Mouth	Nose
Hit Rate	98.3%	98.5%	95.1%	81.2%
False Det.	21.3%	20.8%	42.6%	26.7%

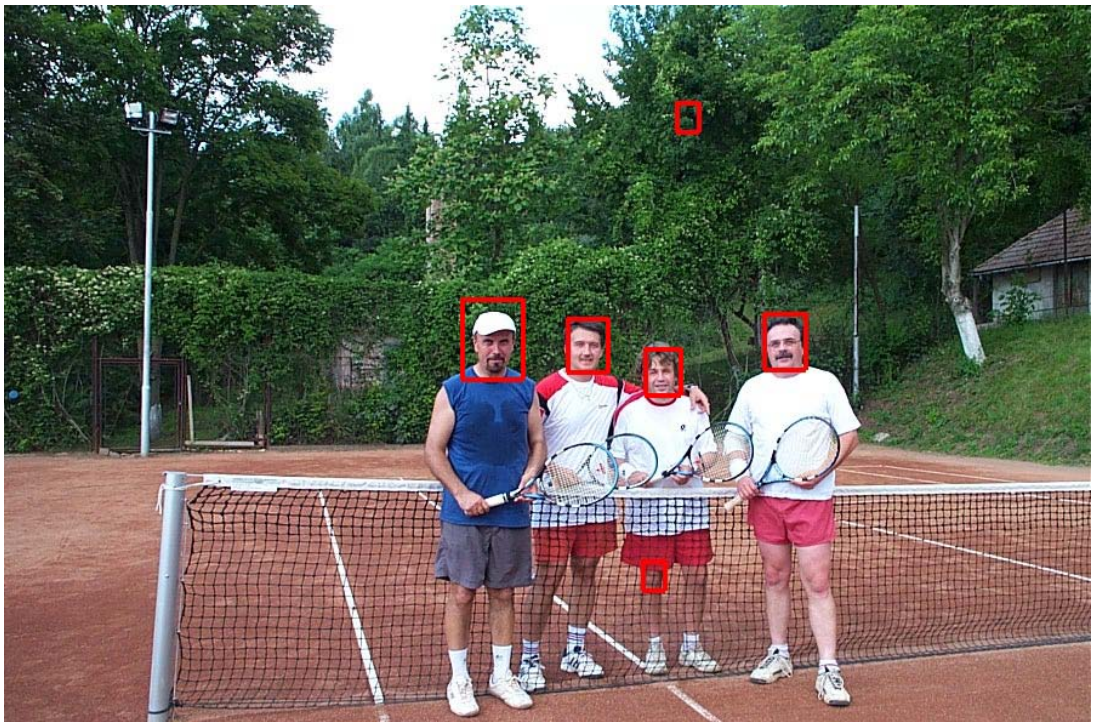


Fig. 5. *Useless false detections*

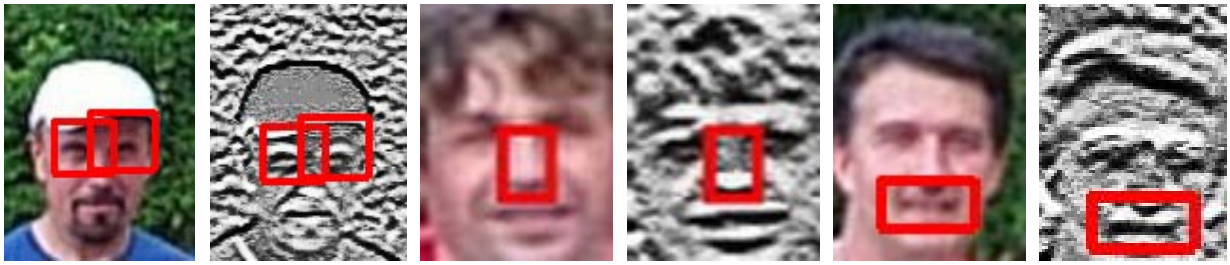
The efficiency of our classifier follows from the measured parameters (see table (4)).
In order to obtain this, our part detector needs patches with enough information. That means it can't decrease below 20×20 pixels.Thus our classifier performs well if the face object exceeds the dimension of 90×120 pixels. Figure (6) shows a few examples from the experimental results.

Table3*. Final stage-classifier*

	False Detection Rate	Hit Rate
Raw patch	14.3%	97%
Wavelet patch	7.8%	98%

Table4*. Final detector*

	Previous (stage 9) performance	New performance
Raw patch	75.5%	98%
No. false det.	150	173



(a) Eyes (b) Eyes (c) Nose (d) Nose (e) Mouth (f) Mouth

Fig. 6. *Example images*

4. Conclusions

In this paper a mixture of classifiers is presented that uses two types of different features. We have proved that the structure information of the target object has an important role in building a robust object classifier.

References

[1]. M. J. P. Viola, "Fast multi-view face detection," Tech. Rep. TR2003-096, Mitsubishi Electric Research Laboratories, Cambridge, July 15, 2003.

[2]. M. J. P. Viola, "Robust real-time object detection," International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154, 2004.

[3]. H. Schneiderman and T. Kanade, "A statistical model for 3d object detection applied to faces and cars," in IEEE Conference on Computer Vision and Pattern Recognition, IEEE, June 2000.

[4]. M. Vidal-Naquet and S. Ullman, "Object recognition with informative features and linear classification," in ICCV, pp. 281–288, 2003.

[5]. B. Leibe and B. Schiele, "Interleaved object categorization and segmentation," in In BMVC, pp. 759–768, 2003.

[6]. R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 264–271, 2003.

[7]. P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," Int. J. Comput. Vision, vol. 61, no. 1, pp. 55–79, 2005.

[8]. V. P. R. Lienhart, A. Kuranov, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection" 25th DAGM Pattern Recognition Symposium, 2003.

[9]. A. K. R. Lienhart, L.Liang, "A detector tree of boosted classifiers for real-time object detection and tracking," Proceedings of International Conference on Multimedia and Expo (ICME'03), vol. 2, pp. 277–280, 2003.

[10]. "http://www.intel.com/research/mrl/research/opencv," 2011

[11]. L. A.-C. J. L.-N. M. Castrillón-Santana, O. Déniz-Suárez, "Face and facial feature detection evaluation - performance evaluation of public domain haar detectors for face and facial feature detection," Proceedings of 3rd International Conference on Computer Vision Theory and Applications (VISAPP'2008) , pp. 167–172, 2008.

[12]. P. I. Wilson and J. Fernandez, "Facial feature detection using haar classifiers," J. Comput. Small Coll., vol. 21, no. 4, pp. 127–133, 2006.

[13]. W. Schulz, M. Enzweiler, and T. Ehlgen, "Pedestrian recognition from a moving catadioptric camera.," in DAGM-Symposium (F. A. Hamprecht, C. Schnörr, and B. Jähne, eds.), vol. 4713 of Lecture Notes in Computer Science, pp. 456–465, Springer, 2007.

[14]. Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, pp. 119-139, Art. No. SS971504, , 1997

[15]. Sz. Lefkovits, "Performance analysis of face detection systems based on haar features," Complexity and Intelligence of the Artificial and Neural Complex Systems, vol. 1, no. 1, 2008.

[16]. Sz. Lefkovits, "Teaching improvements on haar based classifiers," Acta Sapientiae Universitae, vol. 1, no. 1, 2009.

[17]. "Feret database," 2011.

[18]. "Yale database," 2011.

[19]. "Cmu/vacs image data base."

[20]. H. A. S. L. C. Huang, B. Wu, "Omni-directional face detection based on real adaboost," Proceedings of International Conference on Image Processing, vol. 1, pp. 593-596, 2004.