TUNING ASYMBOOST CASCADES IMPROVES FACE DETECTION

I. Visentini, C. Micheloni, G.L. Foresti

University of Udine
Department of Computer Science
Via delle Scienze 206, 33100 Udine, ITALY
visentin, michelon, foresti@dimi.uniud.it

ABSTRACT

The face detection problem is certainly one of the most studied topics in artificial vision. This interest raises from the conscience that this is a crucial step for every system that uses biometric information. Video surveillance and security systems, biometrics, HCI and multimedia applications are some examples of systems that exploit face localization to improve their robustness. AdaBoost and AsymBoost based classifiers are widely used to achieve high performances saving computational time. In this paper, a new reactive strategy to build a strong classifier cascade is provided; at each stage of the cascade a different tradeoff between accuracy and computational complexity is explored. The results will show that this method is effective, and propose a way to construct a rapid and robust multipose detector.

***Index Terms*—** AdaBoost, AsymBoost, Boosting, Reactive learning, Face detection

1. INTRODUCTION

During the last years, the face detection field was deeply investigated. The possible wide application of this technique made grow an increasing interest. In fact, the face detector is a fundamental step in many systems that use biometric information, such as video surveillance, security, Human Computer Interaction, games and multimedia, face and facial expression recognition ones.

When the detection of an object has to be performed on real and complex scenario, the problem is a pattern recognition task. Boosting techniques [1, 2, 3] have been proved to be really efficient in handling the problem especially for what concerns the face detection one. The rising investigation was due to the necessity of a robust detector that cuts off the false negatives (missed faces) by keeping low the false positives (alarms).

AdaBoost [4] was the first implementation of a Boosting algorithm, but unfortunately did not face the problem of the disparity between positive samples (i.e., faces) and negative samples (i.e.,no-faces) that are available in real images. Other considerable developements of the boosting idea are RealAdaBoost [5], in wich the confidence is a real value, FloatBoost [6], that uses an euristic to backward weak hypotheses, and AsymBoost [3], that introduces a mechanism able to asymmetrically weight the two class' samples. The face detection problem applied to a multipose case is presented in [7], where a rapid multiface detector is shown to be effective and capable of handling more pose variations.

The novelty of our solution is represented by a strategy that works on the asymmetry of boosting problems. As consequence of the False Positive (FP) rate achieved so far, the asymmetry parameter is tuned during the learning rounds, allowing an active reaction. We further developed such a strategy to apply it even in context of cascade of classifiers. We will show how, using an automated tuning strategy during both the learning of the single classifiers and the construction of the cascade, at the same time the number of False Negatives can be significantly reduced without affecting the performance of the overall system.

In the remainder of the paper we will give in Section 2 a short description of the boosting techniques, where particulare attention is paid to Adaboost and Asymboost algorithms and their exploiting in cascade. In Section 3, we will resume the new concept of reactive control of the asymmetry during the learning levels. Finally, in Section 5 we will show how the proposed strategy works in multipose case, ensuring good performance and saving computational effort.

2. BOOSTING, ADABOOST AND ASYMBOOST

Boosting algorithms are iterative procedures which produce a linear combination of simple hypotheses $h_1, \ldots, h_T$ to generate a robust ensemble [1]

$$H(x) = sign\left(\sum_{n=1}^{T} \alpha_t h_t(x)\right) \quad (1)$$

where each hypothesis h_i is slighty better than random guessing, α_i are the coefficients used by linear combination.

Such an iterative batch learning algorithm is based on two main ideas:

- a combination of weak hypotheses produces a strong hypothesis;
- a distribution of weights is maintained over the input samples in order to optimally drive the selection of the hypothesis.

Let D be the distribution of weights associated to the samples in the training set. $D_t(i)$ represents the weight of the sample i at the learning round t. At each round, the distribution D_t is modified in order to increase the weight of the misclassified samples, following the formula

$$D_{t+1}(i) = \frac{exp(-y_i \sum_t h_t(x_i))}{\prod_t Z_t} \tag{2}$$

where Z_t is a normalization coefficient.

However, when the training set is highly skewed, as in real world, AdaBoost has some limitations: first of all, the difficulty to minimize false negatives (i.e., the number of misclassified faces). For such a reason, Viola and Jones [3] introduced AsymBoost, or Asymmetric AdaBoost. the weights are updated in a way that the importance of positive samples is increased of a factor $exp(y_i log\sqrt{k})$ at each round

To solve the constraint of a real time computation, more strong classifiers are applied on an image in rapid sequence. At first stages small strong classifiers are applied to reject the majority of the non–faces patterns. In the next levels, more specialized and time consuming strong classifiers are requested to validate the remaining ones. At each level l, the corresponding classifier H_l can validate or reject an input sample; in the first case, the image is given to the next classifier H_{l+1}. In the second case, the sample is definitely rejected.

At the end of the cascade process, the global false alarm ratio FP, defined as the ratio between the number of the misclassified non-faces over the whole number of non-faces, is

$$FP = \prod_l fp_l \tag{3}$$

where fp_l is the ratio at level l. Contrariwise, the overall false negative ratio is the summary of the single levels' ratios fn_l at each level l of the cascade

$$FN = \sum_l fn_l \tag{4}$$

3. REACTIVE LEARNING

The idea to improve the constraints on the FP rate at each level of the cascade and to tune the parameter k for local selection of the weak hypothesis generates a new algorithm. To distinguish this solution from the original algorithm AsymBoost we introduced the notation AsymBoost* [8].

3.1. Flexible false positive rate

The naive algorithm wants the classifier to reach for each level l a value $fp_{l,optimal}$. It can happen that, when the training error decreases very slowly, the algorithm adds many useless features to the classifier without noticing any benefit to the detection rate. For this reason, in the training of the classifier at level l we replaced $fp_{l,optimal}$ with a dynamic threshold, defined as

$$fp'_{l,optimal} = fp_{l,optimal} * \left(\frac{fp_{l,optimal}}{fp_{l-1}}\right) \tag{5}$$

Using this balancing, we can dynamically adapt the false positive threshold to the classifiers' potential.

Moreover, a backtracking strategy has been implemented, to avoid local minima

repeat
  Select the nth feature h_n
until $\sum_{j=p}^{t} fp_{l,j} > fp_{l,t} * (t-p)$

3.2. Balance of asymmetry

The second heuristic refers to the false negatives. Supposing that the FN rate at the level l is quite far from the optimal threshold $fn_{l,optimal}$, at each step t of the training we can assign a different value to $k_{l,t}$, forcing the false negative ratio to decrease when $k_{l,t}$ is high. An effective method to avoid the false positives exponential uncontrolled increment is to assign $k_{l,t}$ depending on the desired value $fp'_{l,optimal}$

$$k_{l,t} \propto \left|(fp'_{l,optimal} - fp_{l,t-1})\right| \in [k_{min}, k_{max}] \tag{6}$$

For each step $t = 0, \ldots, T$, $k_{l,t}$ is supposed to be

$$k_{l,t} = 1 + \frac{fp'_{l,optimal} - fp_{l,t-1}}{fp'_{l,optimal}} \tag{7}$$

Another way to compute k depends on its estimation based on the last two values on the progression of fp_l

$$\hat{k}_{l,t} = 1 + \frac{fp'_{l,optimal} - \hat{f}p_{l,t-1}}{fp'_{l,optimal}} \tag{8}$$

where

$$\hat{f}p_{l,t} = (fp_{l,t-1}) + (fp_{l,t-1} - fp_{l,t-2}) = 2fp_{l,t-1} - fp_{l,t-2}. \tag{9}$$

In this case the prediction has to be corrected *a posteriori* by the real value.

4. OUT–OF–PLANE MULTIPOSE DETECTION

To detect also multiple–pose face patterns, a multiview detection system is presented. The system is based on the coarse–to–fine strategy introduced in [9] and it is similar to the cascade concept. The first levels are learned to detect a generic human face shape, while the subsequent levels detect a more specific pose, as shown in Fig 2.

Each level discards as many negative subwindows as possible, to reduce the computational cost for the next levels. Each level corresponds to an Haar feature-based cascade system, that has to be rapid and quite robust.

Our system is intended only to deal with out–of–plane rotations, in the degrees range $\Theta = [-90, +90]$. The possible degrees of head rotation have been divided into five ranges, as shown in Fig. 1. The full left (right) profile has $\alpha = [-90, -55]$ ([55, 90]), for the left (right) half profile $\alpha = [-55, -20]$ ($[20, 55]$), and the frontal one $[-20, 20]$.

At each range corresponds a set of images which the detector has been trained on. Each level is trained independently of the next one and a bootstrapping system is used to construct the negative set of samples for the training of the following levels. The most probable range of head rotations has the higher *confidence* [2] of all classifiers.

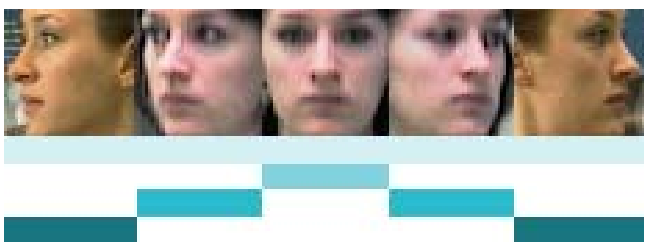

Fig. 1. The range subdivision for head rotations.

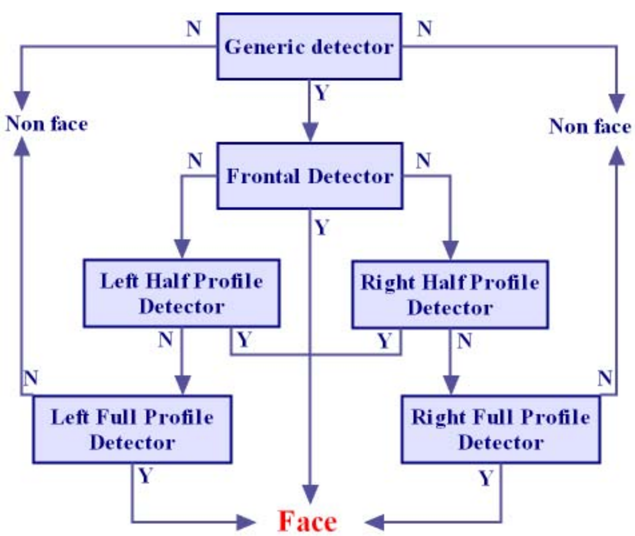

Fig. 2. The coarse–to–fine strategy for the rotation detection scheme is here presented.

5. EXPERIMENTAL RESULTS

5.1. Frontal faces

For the first experiment, a training set consisting on $5,000$ positive (face) samples and $10,000$ negative (non–face) samples, scaled in a standard format 24×24 pixels was used. AsymBoost (that is proved in [3] to be better than AdaBoost) and AsymBoost* (using Eq. (7)) were used to train two Haar-features [2] cascade classifier. In Fig.3, ROC curves for AsymBoost and AsymBoost* on MIT + CMU dataset [10] are presented. In the region of interest, with the same false positive rate, the number of false negatives introduced by AsymBoost* is reduced with the respect to other algorithm.

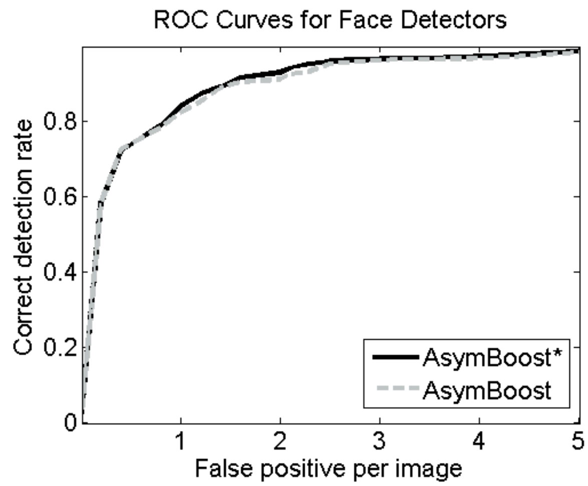

Fig. 3. ROC curves on MIT + CMU dataset for comparison of AsymBoost and AsymBoost*.

The most visible consequence of such an improvement to the asymmetry is the inclusion of less features to the single strong classifier. It means that a lower number of features per level is chosen, thus reducing the overall computational cost, as we already noticed in [8].

5.2. Multiwiew faces

For the second experiment, a set of $8,500$ images cropped at the size 24×24 is divided in frontal and rotated ones and and it is used to train a multipose AsymBoost* based system. Training images are divided following the classification scheme presented in Sec. 4.

The first top level of the pyramid is trained with all face examples. It consists in a three-level cascade, made by 45 features, that rejects about 50 percent of non–faces and has only 0.05 percent of false negatives. At the second level, a second classifier for detecting poses in the range $\Theta = [-55, -20]$ is applied. In the third level, the left profile pose detector ($\Theta = [-90, -55]$) analyses the remaining patterns. The same is symmetrically done for the right profile classifiers.

To test this system we used an internal database of images of 57 different people in 9 different views each. For all the subjects, the position of the eyes is known and the head rotation follows a pre-established position. The ground truths were compared automatically with the detection outputs, and the results are presented in Fig. 4. The detection rate on the y axis is the mean of the correct detection rates of the system on the images corresponding to that degrees subspace on the x axis. The variance, that gives the error margin of the detector, is represented as a lighter area around the mean. As we can see, the detection rate is higher for images with angles close to those used for the training process.

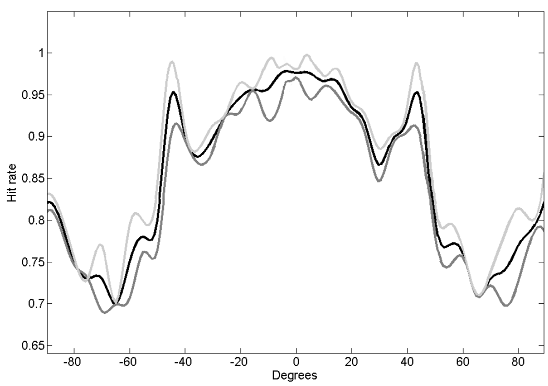

Fig. 4. Detection rate on the multiview image test set.

6. CONCLUSIONS

In this paper we have presented a new reactive learning strategy. Its application to all the stages of the training process results in an improved version of the AsymBoost algorithm. Indeed we have shown that tuning the weight of the asymmetry yields to a smaller false negative value by keeping the false positives within the same range. Besides, the experimental results have shown how the proposed method, applied to well known datasets and to an internal video representing the real scenario, performs better to the standard algorithms as those represented by AdaBoost and AsymBoost.

Acknowledgments

This work was partially supported by the Italian Ministry of University and Scientific Research within the framework of the project "Ambient Intelligence: event analysis, sensor reconfiguration and multimodal interfaces"(2007-2008) and by the European FP6 Project HAMLeT "Hazardous Material Localisation & Person Tracking" (SEC6-SA-204400)

7. REFERENCES

[1] Robert E. Schapire, "Theoretical views of boosting and applications," in *Algorithmic Learning Theory*, Tokyo, Japan, December 1999, vol. 1720, pp. 13–25.

[2] Paul A. Viola and Michael J. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR (1)*, Kauai, HI, USA, 8–14 December 2001, pp. 511–518.

[3] Paul A. Viola and Michael J. Jones, "Fast and robust classification using asymmetric adaboost and a detector cascade," in *NIPS*, Vancouver, British Columbia, Canada, 3–8 December 2001, pp. 1311–1318.

[4] Yoav Freund and Robert E. Schapire, "Experiments with a new boosting algorithm," in *ICML*, Bari, Italy, 3–6 July 1996, pp. 148–156.

[5] Jerome Friedman, Trevor Hastie, , and Robert Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, vol. 38, no. 2, pp. 337– 374, Aprile 2000.

[6] S. Z. Li and Z. Zhang, "Floatboost learning and statistical face detection," *Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1112– 1123, Settembre 2004.

[7] Rong Xiao, Ming-Jing Li, and Hong-Jiang Zhang, "Robust multipose face detection in images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 31–41, Gennaio 2004.

[8] Ingrid Visentini, Christian Micheloni, and Gian Luca Foresti, "Reactive learning strategy for asymboost based face detectors," in *International Conference on Image Analysis and Processing*, Modena, Italy, 10-14 September 2007 2007, p. in press.

[9] ZhenQiu Zhang, MingJing Li, Stan Z. Li, and HongJiang Zhang, "Multi-view face detection with floatboost," in *WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, Washington, DC, USA, 2002, p. 184, IEEE Computer Society.

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