Combining adaBoost with a Hill-Climbing evolutionary feature search for efficient training of performant visual object detector

Yotam Abramson, Fabien Moutarde, Bogdan Stanciulescu, Bruno Steux

▶ **To cite this version:**

Yotam Abramson, Fabien Moutarde, Bogdan Stanciulescu, Bruno Steux. Combining adaBoost with a Hill-Climbing evolutionary feature search for efficient training of performant visual object detector. FLINS2006 on Applied Artificial Intelligence, Aug 2006, Genova, Italy. hal-00435700

HAL Id: hal-00435700
https://hal.science/hal-00435700

Submitted on 24 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMBINING ADABOOST WITH A HILL-CLIMBING EVOLUTIONARY FEATURE SEARCH FOR EFFICIENT TRAINING OF PERFORMANT VISUAL OBJECT DETECTORS

Y. ABRAMSON

Transportation Research Institute,
Technion - Israel Institute of Technology
Haifa 32000, Israel
E-mail: abramson@technion.ac.il

F. MOUTARDE, B. STANCIULESCU AND B. STEUX

Ecole des Mines de Paris, Robotics Laboratory
F-75272 Paris Cedex 06, France
E-mail: {Fabien.Moutarde,Bruno.Steux,Bogdan.Stanciulescu}@ensmp.fr

This paper presents an efficient method for automatic training of performant visual object detectors, and its successful application to training of a back-view car detector. Our method for training detectors is adaBoost applied to a very general family of visual features (called "control-point" features), with a specific feature-selection weak-learner: evo-HC, which is a hybrid of Hill-Climbing and evolutionary-search. Very good results are obtained for the car-detection application: 95% positive car detection rate with less than one false positive per image frame, computed on an independant validation video. It is also shown that our original hybrid evo-HC weak-learner allows to obtain detection performances that are unreachable in reasonable training time with a crude random search. Finally our method seems to be potentially efficient for training detectors of very different kinds of objects, as it was already previously shown to provide state-of-art performance for pedestrian-detection tasks.

1. Introduction

The seminal work of Viola and Jones [1] [2] introduced a new and powerful framework for training of visual object detectors. It is based on the adaBoost algorithm, where each "weak-classifier" assembled in the final strong classifier uses a single and simple image feature. In most works inspired by [1], these features are localized filters similar to Haar wavelet basis. At each adaBoost step, one of them is selected whose weighted error on the

training set is as low as possible.

Following this work, several authors tried different approaches for further improvement of the algorithm, either for speeding up the training, and/or improving the performance of the final detector. Two main directions have been explored: extending or changing completely the set of features, and/or trying to replace the exhaustive search for feature selection by a more efficient process. For instance, McCane and Novins [3] proposed an alternative non-exhaustive search based on a simple "local search" heuristic, and found that they could obtain nearly as good a classifier as in [2], but with a much faster training. Bartlett *et al.* [4] also used some custom heuristic based on initial random selection of a small subset (5%) of all possible features, from which the best one is further refined by some kind of local search (over a set of features obtained from the best initial one by applying various shifting, scaling and reflecting operations). Treptow and Zell [5] proposed to extend the features family proposed by Viola and Jones to a more generalized set of similar features, and to use a specific evolutionary search as "weak learner". They found that adaBoost training with their evolutionary search over their larger feature set produced better detectors than exhaustive search applied to the initial limited feature set.

Simultaneously, we have proposed and tested in [6] and [7] a radically different set of image features: the so-called "control-points" features (see 2.1), and designed a custom "evolutionary" weak-learner for selecting, at each boosting step, an individual feature from the huge control-points feature space. In this paper, we detail our specific weak-learner heuristic, show that it allows to reach detection performances unattainable by crude random search, and present the very good results of our method on a different application: car detection in images from on-vehicle camera (instead of face or pedestrian detection).

2. Experiments

2.1. *Image features*

The very general family of features we use are called "control points" features. This family was first designed and proposed by us in [6], where we already described it. Therefore we hereafter only briefly recall what these features are.

Each feature is defined by two sets of "control points", $\{p_1 \dots p_i\}$ and $\{n_1 \dots n_j\}$, where $i, j \leq K$ (we use $K = 6$ as in [6]), all placed either within the $W \times H$ detection window, or on a half-resolution or quarter

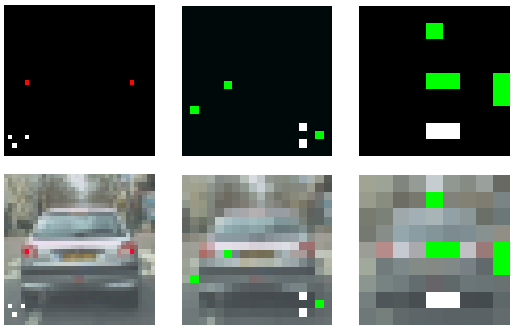

Figure 1. Three examples of control point features: the left one is "full resolution" with 2 "positive" points and 3 "negative" points located in the 36×36 pixels detection window; the middle one is "half-resolution" with 3 positive and 2 negative points testing the 18×18 pixels down-sampled detection window; the last one on the right is "quarter-resolution" with 5 positive and 2 negative points chosen in the 9×9 pixels of the twice down-sampled detection window. The upper row shows the control points by themselves, and the lower row illustrates the application of the same 3 features to a given sub-window extracted from one movie image.

resolution version of the same image. The feature examines the pixel values in $\{p_1 \dots p_i\}$ and $\{n_1 \dots n_j\}$ in the relevant image (full, half or quarter resolution), and answers "yes" if and only if for *every* control-point $p \in \{p_1 \dots p_i\}$ and *every* control point $n \in \{n_1 \dots n_j\}$, $val(p) - val(n) > \theta$ is true, where θ is some *feature-dependant* minimal margin. Note that this last condition is a generalization of the original control-points features defined in [6] (where $\theta = 0$). For the experiments presented here, the detection window size $W \times H$ is 36×36. We refer the reader to [6] for more details, justification and advantages of this family of features.

2.2. *Datasets*

We used data recorded by ourselves with an on-vehicle 320x240 high-end CCD color webcam (LogiTech QuickCam Pro 4000). The training and validation sets were built partly manually (essentially for the positive examples), and partly using the "active learning" approach, with the semi-automated selective sampling implemented by the "SEVILLE" software (see [7] for more details). For creating the positive examples, the sub-image was carefully positioned and sized in order to have the *back* of the car centered, with margins of about 15% around it, as can be checked on positive examples shown on top of figure 2.

Figure 2. Some examples from our training set: various positive car examples on the upper row, and several negative examples on the lower row (note that we define non-car as non-"correctly centered rear of car"). Image examples are all square but of various sizes; they are all warped before training to the same 36×36 detection window, and its 2 sub-sampled versions.

Our final total dataset includes 3291 labeled square images of various sizes, among which 1224 positive examples. This dataset was split in two: 2/3 of randomly chosen examples constituting the 2204 examples (including 791 positive) of our training set, the remaining 1/3 just being used as a validation set to monitor for overtraining. For final testing, in order to ensure an unbiased measure of accuracy, we used an independant recording together with a manually built "ground truth" information. The latter specifies, for each successive image in the movie, the exact position and size of actual "rear of car" positive detections that a perfect detector should output.

2.3. *AdaBoost training*

For detector training, we use the adaBoost algorithm [8], as in [2] and most subsequent papers. AdaBoost requires a "weak learner", i.e. an algorithm which will select and provide, for each adaBoost step, a "good" feature (i.e. with a "low-enough" *weighted* error measured on the training set). The weak learner used by Viola and Jones is just an exhaustive search of all $\approx 180,000$ possible features in their set of features. But in our work, exhaustive search is definitely *not* possible, because we use "control points" features as described in 2.1, and the total number of possible different features is absolutely huge in this family (there are more than 10^{35} of them for our 36×36 detection window size). One of the goals of the present work is precisely to present in detail our weak-learner heuristic specifically designed for efficient exploration of this huge space.

2.4. *Evo-HC hybrid search weak-learner*

The general scheme of our custom-designed hybrid search is the following:

- we start with a first generation of 70 random *simple* features (i.e. with only 2 positive and 2 negative points);
- at each generation:
  (1) we select the 30 best features of the previous generation (those with lowest weighted error on the training set);
  (2) we apply to those 30 best features a hill-climbing consisting in applying to each of them a maximum of 5 *successive* specifically-designed "mutations", each mutation being cancelled if it did not improve the feature;
  (3) we complete the population with 40 new *simple* random features;
- we stop the algorithm when there has been no improvement during 40 consecutive generations, and choose the best feature of the last generation as the "good" selected feature for this adaBoost round.

We call this custom exploration search evo-HC, where "evo" stands for evolutionary, and "HC" for Hill-Climbing. Indeed our search algorithm is largely driven by the "random mutation hill-climbing" step (2) where the best features of the previous generation are further refined by several successive only-improving "mutations". And it is at the same time somehow hybrid with an evolutionary search as for each new generation the 40 worst features (among 70) are discarded and replaced by new random simple features. It should be noted however that only selection (and no crossover) occurs during our evolution. But, as found some time ago by Mitchell *et al.* in [9], Random Mutation Hill-Climbing (RMHC) by itself outperforms Genetic Algorithms in some difficult optimization problems.

The "mutation" operator is specifically designed for the control-points features. It is a random choice between one of the 6 following alterations:

- randomly moving one of the existing control-points within a maximal radius of 5 pixels around its initial position;
- add a new random (positive or negative) control-point;
- remove one of the existing control-points;
- change the working channel (red, blue or green) of the feature;
- change the feature "margin" value θ (see section 2.1) by ± 2;
- change the working resolution (full, half or quarter) of the feature.

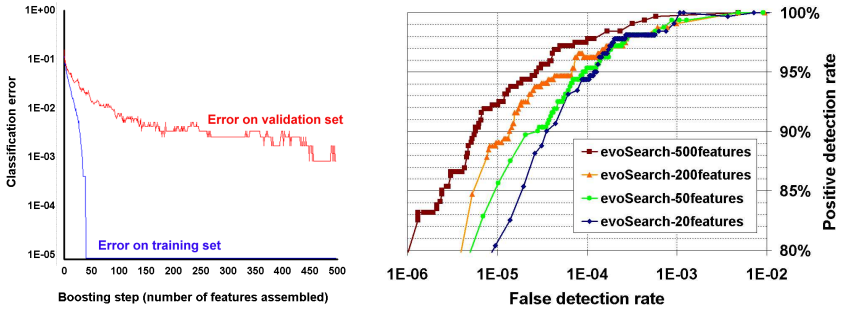

Figure 3. Monitoring of adaBoost training using our evo-HC search: on the left, typical training curve during training; on the right, evolution of ROC curve during training (generally moving left and up along boosting iterations).

3. Results

As is customary for assessment of detectors (*cf.* for instance [2]), most of our evaluations and comparisons are based on the "Receiver Operating Characteristic" (ROC) curves of detectors. These ROC curves are graphical plots of sensitivity *vs.* specificity (here, positive detection rate *vs.* false detection rate) as the discrimination threshold is varied (see *e.g.* [10]).

3.1. *Detection performance with our evo-HC weak-learner*

Our training method works very well for the car detection problem, as illustrated by the very good ROC curve obtained for the 500-features detector (see upper curve on right of figure 3). The positive detection rate for this classifier is $\approx 95\%$ for a false detection rate of $1 : 40,000$. The latter rate approximately corresponds to one false detection per image frame in the camera stream, as in our setup $33,227$ sub-windows of various locations and sizes are tested in each image.

The typical computation time for adaBoost training on our 2204 car/non-cars examples set using our evo-HC search weak-learner is ≈ 28 minutes (on a Pentium IV 3Ghz) for each 100 boosting cycles (and thus for each 100 features added). The total training time for a 500-features detector is therefore ≈ 2.5 hours.

3.2. *Comparing evo-HC to "random search" weak-learners*

In order to assess the exploration power of our evo-HC search weak-learner, we conducted a systematic comparison with an alternative trivial weak-learner: a simple random search. The latter is extremely simple: for each adaBoost step, we simply try successively $maxTrials$ randomly generated features. Note that, contrary to what we do in our evo-HC, the feature

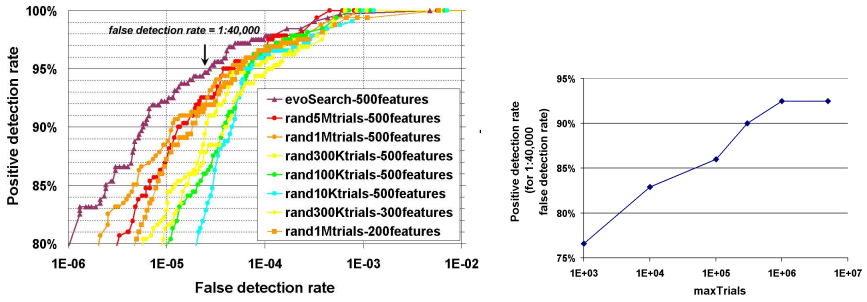

Figure 4. Comparison of detection performance attainable with our evo-HC search, to that of detectors obtained using random search: left, the ROC curves of the latter are always below, when comparing detectors with the same or a smaller number of features; right, slow improvement of random-search detectors with increasing $maxTrials$, reaching a maximum *before* the performance obtained with our evo-HC search.

randomization procedure naturally covers *all possible features*. The only parameter of the random search weak-learner is $maxTrials$, the number of randomly generated features for each adaBoost step.

It is clear from ROC curves on figure 4 that the detection performance of the best detector using our evo-HC search weak-learner is *definitely higher* than that of any of the detectors we obtained using random search weak-learner. This is true not only for the most complex detector (i.e. with 500-features), but also for any given maximal number of features allowed for the detector. Also, even though some detectors obtained with random search with very high value of $maxTrials$ produce "lower but acceptable" detection performance, it should be noted that the corresponding training time was already much higher (nearly 10 times) than the training time with our evo-HC weak-learner.

4. Conclusions and discussion

In this paper, we have shown that adaBoost training with control-points features and a specially-designed weak-learner (evo-HC, a hybrid of Hill-Climbing and evolutionary search) can produce a very good "car detector" reaching 95% positive detection rate for $1 : 40,000$ false detection rate (i.e. less than 1 false alarm per video frame).

Moreover, we have conducted a series of tests to compare these good results to what can be obtained using an alternative very simple "random search" weak-learner. It was shown that it is apparently *not* possible (even if increasing the number of tested features to values implying unreasonable training times) to get as good a final classifier as the one that was obtained using our evo-HC search weak-learner. It should also be noted that the

training time of a 500-features detector on our $\approx$ 2000 images training set is only $\approx$ 2.5 hours on a typical desktop. A deeper study and comparison of our weak-learner with other *sophisticated* search algorithms should now be conducted. Nevertheless, the evidences presented here already indicate that our evo-HC search is indeed a quite efficient weak-learner, with good exploratory power, allowing performant selection of several hundreds of discriminative features among the $\approx 10^{35}$ possible "control-points" features.

Finally, as we had previously successfully applied our method for training an acceptable face-detector [6], and a state-of-the-art pedestrian-detector [7], it seems that our adaBoost training with evo-HC weak-learner exploring "control-points" feature space can be an efficient and *general* method for training visual detectors of any kind of objects.

References

1. P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR'01)*, Kauai , Hawai, USA (2001).
2. P. Viola and M. Jones, Robust real-time object detection. *International Journal of Computer Vision* **57(2)** (2004).
3. B. McCane and K. Novins, On training cascade face detectors. In *Image and Vison Computing*, pages 239-244, New Zealand (2003).
4. M.S. Bartlett, G. Littlewort, I. Fasel and J.R. Movellan, Real time face detection and facial expression recognition: Development and application to humn-computer interaction. In *CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer interaction*, Vancouver (2003).
5. A. Treptow and A. Zell, Combining adaBoost learning and evolutionary search to select features for real-time object detection. In *Proc. of IEEE Congress on Evolutionary Computation (CEC 2004)*, Portland, Oregon, vol. 2, pp. 2107-2113, IEEE Press (2004).
6. Y. Abramson, B. Steux and H. Ghorayeb, YEF real-time object detection. In *Proc. of Intl. Workshop on Automatic Learning and Real-Time (ALART'05)*, Siegen, Germany (2005).
7. Y. Abramson and Y. Freund, SEmi-automatic VIsuaL LEarning (SEVILLE): a tutorial on active learning for visual object recognition. In *Proc. of IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego (2005)
8. Y. Freund and R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory: EuroCOLT'95*, pages 23-37, Springer-Verlag (1995).
9. M. Mitchell, J.H. Holland and S. Forrest, When will a genetic algorithm outperform Hill-Climbing ? In *Advances in Neural Information Processing Systems*, vol. 6, pages 51-58, Morgan Kaufmann Publishers Inc. (1994).
10. J.A. Hanley, Receiver operating characteristic (ROC) methodology: The state of the art. *Critical Reviews in Diagnostic Imaging*, **29**:307-335 (1989).