České vysoké učení technické v Praze,
Fakulta elektrotechnická

Czech Technical University in Prague,
Faculty of Electrical Engineering

Doc. Dr. Ing. Jiří Matas

Sekvenční rozhodování v problémech (počítačového vidění) s časovým omezením.

Sequential Decision Making for Time-Constrained (Computer Vision) Problems.

Summary:

In many computer vision decision making problems, e.g. detection or classification, both the error rates and the evaluation time characterise the quality of a solution. Yet none of the currently popular learning algorithms such as support vector machines, neural networks or AdaBoost that are typically employed to train such a decision-making system optimise (or even consider) the evaluation time explicitly.

The trade-off between decision quality in terms of error rates and the time-to-decision expressed by the number of measurements has been studied in the 1940's by A. Wald. In his theory of sequential decision-making, Wald proved that the optimal sequential strategy in terms of the shortest average time to decision (number of measurements used) given predefined error rates is the sequential probability ratio test (SPRT). However, Wald's theory is based on measurements that are assumed to be selected and ordered *a priori*. Moreover, it is assumed that either the measurements are class-conditionally independent or their joint probability density functions are known.

We show how these limitation can be overcome by selecting the relevant measurements by AdaBoost. The joint conditional density of all measurements, whose estimation is computationally intractable, is approximated by the class-conditional response of the sequence of strong classifiers. The choice is justified by the asymptotic properties of AdaBoost-trained strong classifiers. The resulting algorithm, called WaldBoost, integrates AdaBoost-based measurement selection and Wald's optimal sequential probability ratio test.

As an application of WaldBoost, we demonstrate how existing (slow) binary decision algorithms can be approximated by a (fast) trained WaldBoost classifier. The WaldBoost learning is used to minimise the decision time of the emulated algorithm while guaranteeing predefined approximation precision. Moreover, the WaldBoost algorithm together with bootstrapping is able to efficiently handle an effectively unlimited number of training examples provided by the implementation of the approximated algorithm.

Two interest point detectors, the Hessian-Laplace and the Kadir-Brady saliency detectors, are emulated to demonstrate the approach. Experiments show that while the repeatability and matching scores are similar for the original and emulated algorithms, a 9-fold speed-up for the Hessian-Laplace detector and a 142-fold speed-up for the Kadir-Brady detector is achieved.

Souhrn:

V mnoha rozhodovacích problémech v oblasti počítačového vidění (např. u detekce nebo klasifikace) charakterizuje kvalitu procesu rozhodování nejen jeho chybovost, ale také doba výpočtu nutná pro rozhodnutí. Nicméně žádný z běžných učících se algoritmů jako jsou neuronové sítě, metody SVM nebo Adaboost, které se používají pro učení těchto rozhodovacích systémů, neoptimalizuje, dokonce ani neuvažuje, dobu nutnou pro rozhodnutí.

Kompromis mezi chybovostí rozhodnutí a dobou rozhodování vyjádřenou počtem měření byla zkoumána v čtyřicátých letech dvacátého století A. Waldem. Ve své teorii sekvenčního rozhodování Wald dokázal, že optimální sekvenční strategie, t.j. strategie s nejkratším průměrnou dobou rozhodnutí (reprezentovanou počtem měření), která splní omezení na chybovost, je tzv. Waldův sekvenční test založený na věrohodnostním poměru. Omezená aplikovatelnost Waldovy teorie je dána předpokladem, že je a priori známo pořadí jednotlivých měření a že měření jsou nezávislá nebo jejich sdružené hustoty pravděpodobnosti jsou známy.

Přednáška ukazuje, jak lze tato omezení překonat tím, že výběr a pořadí měření uskutečníme pomocí metody strojového učení AdaBoost a nahradíme odhad vysokorozměrné sdružené hustoty všech měření, který je prakticky výpočetně nerealizovatelný, posloupností odezev tzv. silného klasifikátoru. Tato aproximace je opřena o asymptotické vlastnosti silného klasifikátoru učeného metodou AdaBoost. Výsledná metoda, nazvaná WaldBoost, integruje algoritmus AdaBoost a Waldův optimální sekvenční test založený na věrohodnostním poměru.

Možnosti metody WaldBoost demonstrujeme na aplikaci, ve které jsou existující (pomalé) binární rozhodovací algoritmy aproximovány (rychlým) klasifikátorem automaticky naučeným algoritmem WaldBoost. WaldBoost učení minimalizuje dobu rozhodování vybraných detektorů při zaručené kvalitě aproximace. V úloze se projeví výhoda algoritmu WaldBoost, který je pomocí tzv. bootstrappingu schopen efektivně zpracovávat téměř neomezený počet trénovacích příkladů, které jsou získany spouštěním aproximovaného algoritmu.

Přistup je ověřen na dvou detektorech bodů zájmu, tzv. ”Hessian-Laplace ” a ”Kadir-Brady saliency”. Výsledné emulátory mají shodnou opakovatelnost a kvalitu lícování jako původní algoritmus, ale emulátor dosáhl 9-násobné zrychlení u detektoru Hessian-Laplace a 142-násobné zrychlení u detektor Kadir-Brady.

Keywords: sequential decision making, boosting, AdaBoost, computer vision, machine learning, feature point detection.

Klíčová slova: sekvenční rozhodování, boosting, AdaBoost, počítačové vidění, strojové učení, detekce bodů zájmu.

Contents

1. Introduction

Standard learning algorithms like support vector machines, AdaBoost or neural networks are designed primarily with the objective of error minimisation and generalisation to unseen data; small training size performance is a common important concern. Typical evaluation methodology for learning algorithms reflects this focus – measures like error rates, the precision-recall curve or the false positive and the false negative rates are usually reported.

However, in many practical decision problems, another aspect of the trained classifier becomes critical – *the time-to-decision*. Very few approaches consider time-to-decision as an integral part of the learning task. We present the WaldBoost learning algorithm, introduced by us in [30], which handles the precision-speed trade-off automatically and produces a quasi-optimal *sequential classifier* minimising the decision time while guaranteeing predefined error rates.

Learning to be fast. The history of the formulation of a classification task with time-to-decision vs. precision trade-off dates back to Wald's sequential analysis [34]. Wald posed the problem as a constrained optimisation and found a quasi-optimal solution to it – the sequential probability ratio test (SPRT). Wald also showed that on average, a fewer number of observations is needed to achieve the required error rates in comparison with a decision strategy making a fixed number of measurements. However, Wald's theory assumes knowledge of the class conditional probabilities and it does not consider learning and estimation issues. The theory of sequential decision-making has been further developed and enriched [29] and is now used as a basic and well known tool in statistics.

In 1987, Rivest [26] studied learnability of decision lists (which could be seen as sequential classifiers) in the context of Boolean functions but without optimising the evaluation time. Baker and Nayar [1] looked at the problem of efficiency of classification in the context of multi-class classification, where the task is to effectively distinguish one class out of many. To this end, they developed a theory of pattern rejectors which can be interpreted as sequential classifiers in the class space. A practical learning approach to the time-to-decision vs. precision trade-off has been proposed by Viola and Jones [33], who build an ordered set of increasingly complex classifiers that were applied sequentially to a progressively smaller faction of the data. The "classifier cascade" method requires the user to define the complexities of individual classifiers and does not optimise the time vs. precision trade-off directly. Consequently, many variations on the method have appeared in the literature [35, 12, 4]. The SoftCascade [3] algorithm presents a systematic but a rather brute force approach to the precision vs. speed optimisation problem.

Wald's sequential decisions are based on measurements that are assumed to be selected and ordered *a priori*. Moreover, it is assumed that either the measurements are class-conditionally independent or their joint probability density functions are known. We show how this limitation can be overcome by selecting the relevant measurements by AdaBoost[28]. The joint conditional density of all measurements, whose estimation is computationally intractable, is approximated by the class-conditional response of the sequence of strong classifiers. The choice is justified by the asymptotic properties of

AdaBoost-trained strong classifiers. The WaldBoost algorithm integrates AdaBoost-based measurement selection and Wald's optimal sequential probability ratio test.

The WaldBoost approach has successfully been applied to a number of problems: face detection [30], on-line tracking [9], and emulation of feature detectors [31]; an overview of computer vision methods based on the WaldBoost can be found in [19].

Besides introducing the WaldBoost algorithm, its application to the problem of feature detector emulation is presented. We chose this application since the problem of emulation with a learned system is a novel and general idea, applicable outside the field of computer vision.

2. WaldBoost

WaldBoost [30] is a greedy learning algorithm which finds a quasi-optimal sequential strategy minimising the average evaluation time while preserving required quality of the decision for a given binary-valued decision problem. More formally, WaldBoost finds a *sequential decision strategy* S^* such that

$$S^* = \arg\min_S \bar{T}_S \qquad \text{subject to} \quad \beta_S \leq \beta, \quad \alpha_S \leq \alpha \tag{1}$$

for specified α and β. $\bar{T}_S$ is average time-to-decision expressed in the number of measurements evaluated[1], α_S is false negative and β_S false positive rate of a sequential strategy S.

A sequential decision strategy is a sequence of decision functions $S = (S_1, S_2, \ldots)$ where $S_t : (x_1, \ldots, x_t) \rightarrow \{-1, +1, \sharp\}$. The strategy S takes one more measurement, x_t, at a time and in step t makes a decision S_t based on $(x_1, \ldots, x_t)$. The '$\sharp$' sign stands for a "continue" (still undecided) decision. If a decision is '$\sharp$', x_{t+1} is measured and S_{t+1} is evaluated. Otherwise, the output of S is the class returned by S_t.

To find the optimal sequential strategy S^* to the problem (1), the WaldBoost algorithm employs the AdaBoost algorithm [28] for measurement selection and Wald's sequential probability ratio test (SPRT) [34] for finding thresholds which are used for decision-making. The SPRT, a quasi-optimal solution of the problem (1), is very simple – in each step, the likelihood ratio is compared with a fixed threshold. SPRT is a sequential strategy S^* defined as:

$$S_t^* = \begin{cases} +1, & R_t \leq B \\ -1, & R_t \geq A \\ \sharp, & B < R_t < A \end{cases} \tag{2}$$

where R_t is the likelihood ratio

$$R_t = \frac{p(x_1, ..., x_t | y = -1)}{p(x_1, ..., x_t | y = +1)} \,. \tag{3}$$

The constants A and B are set according to the required error of the first kind α and error of the second kind β. Optimal A and B are difficult to compute in practice, but tight bounds are easily derived.

---

[1] In many cases, the time (cost) required to obtain different measurements varies significantly. In such cases, the number of measurements is not a good surrogate for time-to-decision. The generalization of WaldBoost to measurements with different costs is thus of interest.

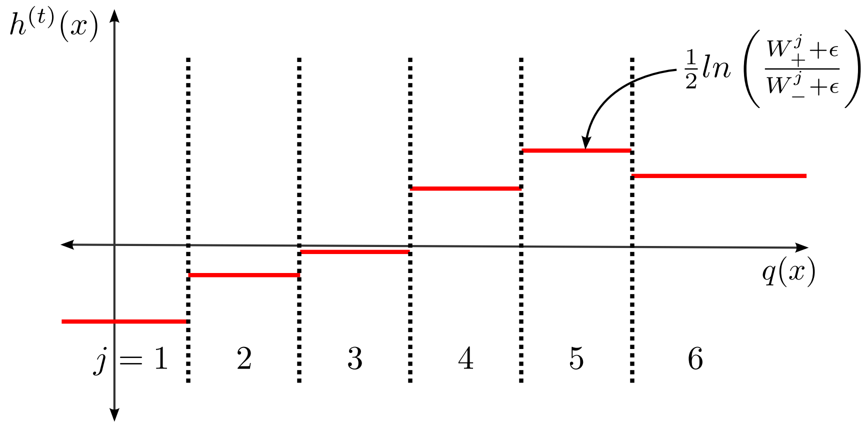


FIGURE 1. The domain-partitioning weak classifier. The response of feature $q(x)$ on object x is partitioned into bins $j = 1, \ldots, K$. The leftmost and the rightmost bins cover respective half-spaces. In each bin j, the response of the weak classifier $h(x)$ is computed from the sum of positive (W_+^j) and negative (W_-^j) weights of training samples falling into the bin. To avoid numerical problems, a smoothing factor ϵ is used [28].

Such test is easy to evaluate for i.i.d. measurements where the likelihood ratio is easy to estimate. However, when the measurements are not i.i.d., the likelihood ratio estimation easily becomes intractable and the ordering of the measurements has to be specified. To overcome these problems, WaldBoost uses the AdaBoost algorithm as a measurement selector and also for projecting measurements (see equation 4) to a 1D subspace where likelihood ratio estimation is tractable. This is justified by the fact that the response of AdaBoost, $f_t(x)$, converges to the likelihood ratio [7].

The AdaBoost algorithm greedily selects *weak classifiers* $h^{(t)} : \mathcal{X} \rightarrow \mathbb{R}$ which are combined linearly into a strong classifier

$$f_T(x) = \sum_{t=1}^{T} h^{(t)}(x). \tag{4}$$

The domain-partitioning weak classifiers [28] are used, each one based on a single (visual) feature (see Figure 1). The response of the weak classifiers found by the AdaBoost algorithm are used as measurements for the sequential strategy in the WaldBoost algorithm.

The input of the learning algorithm (Algorithm 1) is a pool of positive and negative samples $\mathcal{P}$, a set of features $\mathcal{F}$ - the building blocks of the classifier, and the bounds on the final false negative rate, α, and the false positive rate, β. The output is an ordered set of weak classifiers $h^{(t)}$, $t \in \{1, \ldots, T\}$ (i.e. measurements) and a set of SPRT thresholds $\theta_A^{(t)}$, $\theta_B^{(t)}$ optimising (1) for all lengths $t = 1, \ldots, T$. The thresholds are applied directly

Algorithm 1 WaldBoost Learning

Input:
- sample pool $\mathcal{P} = \{(x_1, y_1), ..., (x_m, y_m)\}$; $x_i \in \mathcal{X}, y_i \in \{-1, 1\}$,
- set of features $\mathcal{F} = \{q_s\}$,
- desired final false negative rate α and false positive rate β,
- the number of iterations T.

Sample randomly the initial training set $\mathcal{T}$ from the pool $\mathcal{P}$

For $t = 1, ..., T$

1. Find $h^{(t)}$ by AdaBoost using $\mathcal{F}$ and $\mathcal{T}$ and add it to the strong classifier
$$f_t(x) = \sum_{r=1}^{t} h^{(r)}(x)$$
2. Find decision thresholds $\theta_A^{(t)}$ and $\theta_B^{(t)}$ for f_t using $\mathcal{P}$
3. Bootstrap: update the sample pool $\mathcal{P}$ and sample a new training set $\mathcal{T}$

end

Output: ordered set of weak classifiers $h^{(t)}$ and
thresholds $\theta_A^{(t)}$ and $\theta_B^{(t)}$.

to the strong classifier response f_t (not to the likelihood ration as in SPRT) and are set to $\pm\infty$ if no threshold was found in some learning step.

In the learning stage, the selection of a weak classifier is by far the most time consuming operation. To keep the speed and memory requirements of the training process acceptable, a subset $\mathcal{T}$ is sampled out of the large sample pool $\mathcal{P}$; the selection of the best weak classifier is based on $\mathcal{T}$. The SPRT thresholds are efficiently computed on the whole pool.

The sequential nature of the WaldBoost classifier also affects the sample pool and the training set during the learning. In each round, the already decidable samples in the pool (see explanation for WaldBoost evaluation below) are removed from the learning process and a new training set $\mathcal{T}$ is sampled from the reduced pool.

During evaluation of the classifier (Algorithm 2) on a new sample x, one weak classifier is evaluated at time t and its response is added to the strong classifier response function f_t. It is then compared to the corresponding thresholds and the sample is either classified as positive or negative, or the next weak classifier is evaluated and the process continues
$$S_t(x) = \begin{cases} +1, & f_t(x) \geq \theta_B^{(t)} \\ -1, & f_t(x) \leq \theta_A^{(t)} \\ \text{continue}, & \theta_A^{(t)} < f_t(x) < \theta_B^{(t)} \ . \end{cases} \tag{5}$$

If a sample x is not classified even after evaluation of the last weak classifier, a user defined threshold γ is imposed on the real-valued response $f_T(x)$.

In our interest point detection application of WaldBoost, an arbitrary number of both positive and negative samples is available for bootstrapping. However, when positive

Algorithm 2 WaldBoost Classification

Given: $h^{(t)}, \theta_A^{(t)}, \theta_B^{(t)}, \gamma \quad (t = 1, \dots, T)$
Input: a classified object x.
For $t = 1, \dots, T$
If $f_t(x) \geq \theta_B^{(t)}$, classify x to the class $+1$ and terminate
If $f_t(x) \leq \theta_A^{(t)}$, classify x to the class -1 and terminate
end
If $f_T(x) > \gamma$, then classify x as $+1$ else classify x as -1.

samples were bootstrapped, i.e. early positive classification was allowed in equation (5), all early positive decisions had confidence close to $\theta_B^{(t)}$ and precise localisation via the non-maximum suppression algorithm (see Section 4) was not possible. Thus, we adopted the same asymmetric version of WaldBoost as used in [30], i.e. setting β to zero. The strategy becomes

$$S_t(x) = \begin{cases} -1, & f_t(x) \leq \theta_A^{(t)} \\ \text{continue}, & \theta_A^{(t)} < f_t(x) \end{cases} \tag{6}$$

and only decisions for the negative class are made early during the sequential evaluation of the classifier. A (rare) positive decision can only be reached after evaluating all T classifiers in the ensemble. For problems where the non-maximum suppression algorithm is not applied, the strategy (5) can be used directly.

3. Emulating a binary-valued black box algorithm with WaldBoost

The main idea of the this application of WaldBoost is to look at an existing algorithm as a black box performing some useful binary decision task. The black box algorithm is run on a large dataset of images which provides almost unlimited number of training samples which are used to train a sequential classifier emulating the black box algorithm behaviour. The user's optimisation effort is thus transformed into a much simpler task of finding a suitable set of features which are used in the WaldBoost training.

The main components of the proposed learning system are shown in Figure 2. The black box algorithm provides positive and negative outputs that form a labelled training set. The WaldBoost learning algorithm (see Section 2) builds a classifier sequentially and when new training samples are needed, it bootstraps the training set by running the black box algorithm on new images. Only the samples undecided by the current classifier are used for further training. The result of the process is a WaldBoost sequential classifier which emulates the original black box algorithm.

The training loop uses the fact that the black box algorithm can provide practically unlimited number of labelled training samples. Note that this is in contrast to commonly used human labelled data which are difficult to obtain. The bootstrapping technique [32] is used to effectively update the training set.

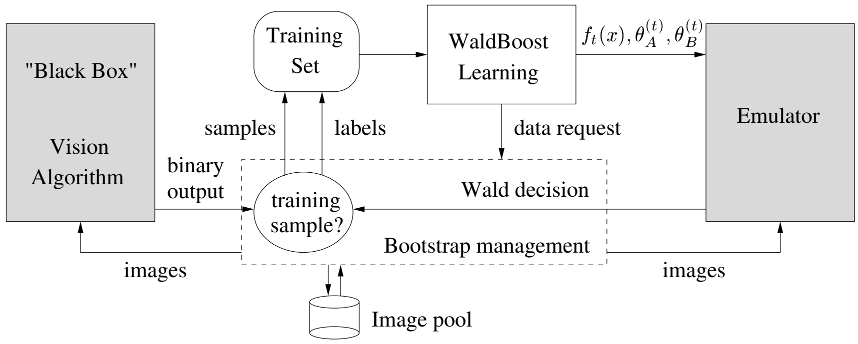


FIGURE 2. The proposed learning scheme.

In the context of fast black box algorithm emulation, what distinguishes training for different algorithms is the feature set F. A suitable set has to be found for every emulated algorithm. The set F can be very large and does not need to be homogeneous. It may contain Haar-like features [33], LBP [25, 8], histograms of gradients, etc. The WaldBoost algorithm selects a suitable subset while optimising the time-to-decision. WaldBoost minimises the average number of evaluated measurements which is the same as minimisation of time-to-decision only when computational complexity of the different types of features is (roughly) the same. The condition is satisfied by the feature set F adopted in the experiments.

Learning interest point detectors. There has been much work on the general interest point detection problem [24]. To our knowledge, learning techniques have been applied only to the parameter tuning, not to the whole process of interest point detector design. Lepetit and Fua [16] treated interest points matching as a classification problem, learning the descriptor. Rosten and Drummond [27] used learning techniques to find parameters of a hand-designed tree-based Harris-like corner classifier. Their motivation was to speed-up the detection process, but the approach is limited to the Harris corner detection. Martin et al. [18] learnt a classifier for edge detection, but without considering the decision time and with significant manual tuning. They tested a number of classifier types with the conclusion that a boosted classifier was comparable in performance to other classifiers and was preferable for its low model complexity and low computational cost.

The most closely related approach to our method is that of Dollár et al. [5] who use learning techniques to train an edge detector. The paper shows impressive examples of applications of such detector. Nevertheless, Dollár et al. were primarily concerned with the accuracy of the detector and did not consider speed. There has also been high interest in speeding up various interest point detectors manually, i.e. without training. Grabner et al. proposed a fast version of the SIFT detector [10] and Bay et al. proposed a fast approximation-based interest point detector called SURF [2].

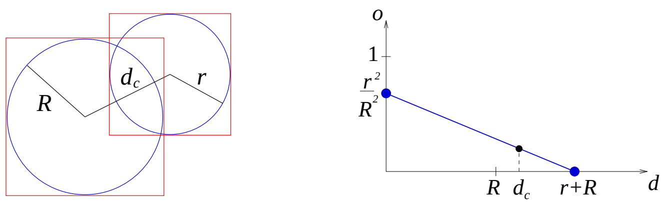


FIGURE 3. Overlap definition for the non-maximum suppression scheme. For details, see the text.

4. Emulated scale invariant interest point detectors

In order to demonstrate the approach, two similarity-invariant interest point detectors have been chosen: (i) Hessian-Laplace [23] detector, which is a state of the art similarity-invariant detector, and (ii) Kadir-Brady [14] saliency detector, which has been found valuable for categorisation, but is about $100\times$ slower than the Hessian-Laplace detector. Binaries of both detectors were downloaded from the web page [21]. We follow standard test protocols for evaluation as described in [24]. Both detectors are similarity-invariant (not affine), which can be easily implemented by running a sequential test at each position and scale in the scanning window approach [33].

For both detectors, the set F includes the Haar-like features proposed by Viola and Jones [33], plus a centre-surround feature from [17], which has been shown to be useful for blob-like structure detectors [10]. Haar-like features were chosen for their high evaluation speed (due to integral image representation) and because they have a potential to emulate the Hessian-Laplace detections [10]. The only difference to the original Viola and Jones feature set is that the feature response is not normalised by a window standard deviation since the intensity contrast is important for both Hessian-Laplace and Kadir-Brady detectors.

For the entropy-based Kadir-Brady saliency detector emulation, however, the Haar-like features were not sufficiently accurate. To overcome this we introduced "variance" features based on the integral images of squared intensities. They are computed as an intensity variance in a given rectangle.

An essential part of a detector is the **non-maximum suppression** algorithm. Here the input to the non-maximum suppression differs from that obtained in the original detectors. Instead of having a real-valued feature response over whole image, sparse responses are returned by the WaldBoost detector. The accepted positions get the real-valued confidence value f_T, but the rejected positions have the "confidence" f_t around the $\theta_A^{(t)}$ value depending on the time t when they have been rejected. These values are incomparable, thus a typical quadratic interpolation and a local maximum search cannot be applied. Instead, the following algorithm is used.

Any two detections are grouped together if their overlap is higher than a given threshold (parameter of the application). Only the detection with maximal f_T in each group is preserved. The overlap computation is schematically shown in Figure 3. Each detection is represented by a circle inscribed to the corresponding scanning window (Figure 3, left). For two such circles, let us denote the radius of the smaller circle as r, the radius of the bigger one as R, and the distance of the circle centres as d_c. Exact overlap can be easily computed in two cases. First, when the circle centres coincide, the overlap is $o = r^2/R^2$. It equals to one for two circles of the same radius and decreases as the radiuses become different. Second, when two circles have just one point in common ($d_c = r+R$), the overlap is zero. These two situations are marked by blue dots in Figure 3, right. Linear interpolation (blue solid line in Figure 3, right)

$$o = \frac{r^2}{R^2}\left(1 - \frac{d_c}{r+R}\right) \tag{7}$$

is used to approximate the overlap between these two states.

5. Results

Two detectors are emulated in the experiments: Hessian-Laplace [23] and Kadir-Brady [14] saliency detector. The Hessian-Laplace is a state-of-the-art detector of blob-like structures used in many applications. Its simplicity allows transparent analysis of obtained results. The Kadir-Brady detector incorporates entropy measure to find salient regions. It shows rather poor results in classical repeatability tests [24] but has been successfully used in several recognition tasks [6, 36]. However, its main weakness for practical applications is its very long computation time in order of minutes per image. Standard versions of the detectors provided by their authors were downloaded from the interest point detection web page [21].

To collect positive and negative samples for training, an emulated detector is run on a set of images of various sizes and content (nature, urban environment, hand drawn, etc.). To create the sample pool we used 1300 images randomly chosen from the non-skin image database introduced in [13]. The detector assigns a scale to each detected point. Square patches of the size twice the scale were used as positive samples. Negative samples were collected from the same images at positions and scales not covered by positive samples.

The size of the training set $\mathcal{T}$ was 10,000 (half positive and half negative samples) in all experiments. The training set was sampled from the pool $\mathcal{P}$ by the quasi-random weighted sampling + trimming method (QWS+) [15]. The QWS+ sampling has been shown to reduce the variance of hypothesis error estimate and to improve the classifier performance compared to other sampling strategies. Moreover, with QWS+ sampling, AdaBoost performance becomes relatively insensitive to the training set size.

5.1. Hessian-Laplace emulation

The Hessian-Laplace detector was used with threshold 1000 to generate the training set. The value was empirically chosen to achieve similar number of detections as in [24]. The same threshold was used throughout all the experiments for both learning and evaluation.

The detector has been assessed in standard tests proposed by Mikolajczyk et al. [24]. The ground truth is given by a homography between the first and the other images in the sequence. The tests are based on two measures: (i) the repeatability measure, (ii) the matching score.

(i) Repeatability measure. To assess the quality of an interest point detector in varying acquisition conditions of the same scene the repeatability measure is used [24]. The measure is defined for two sets of elliptical regions – one set for one image. It is computed as *the ratio between the number of region-to-region correspondences and the smaller of the number of regions in the pair of images*. The mutual correspondence of two regions is claimed when the overlap error is smaller than some threshold. The measure takes into account several other technical issues such as uniqueness of matches and is fully defined by a Matlab script [21]. In all experiments, the overlap error threshold is fixed to 40 % as in most of the experiments in [24].

(ii) The matching score test aims at predicting performance of the detectors in matching and correspondence finding applications. The matching score, defined in [24], is the number of correct matches divided by the smaller number of correspondences in the common part of the two images. A pair of elliptical regions is counted as a *correct match* if (1) their overlap error is smaller than 40 %, and (2) their descriptors are sufficiently similar (for details, see [24]).

Selection of the false negative rate α. The value of α balances the trade-off between WaldBoost detector speed and precision. Figure 4a-c shows performance of the detector for several α values on the BOAT sequence. The value of α also significantly influences the number of detections before the final thresholding by γ (Figure 4d).

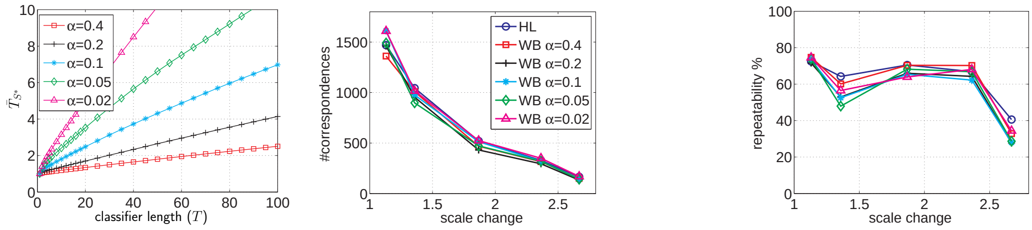
For a certain range of α values, it is possible to set the final threshold γ (Algorithm 2) to reach the number of correspondences similar to that of the emulated detector (Figure 4b). With such threshold γ, the repeatability and the number of correct correspondences is almost identical for all tested values of α throughout the test sequence (Figure 4c).

Increasing α leads to faster evaluation (Figure 4a) but also to less detections (Figure 4d) before imposing the final threshold γ. In some applications it may be useful to produce more detections by changing the γ threshold.

Similarly to the original detector, the WaldBoost emulator imposes a threshold on the classifier response. We set α to 0.2 as a compromise: the classifier is already very fast (see Table 1) and yet the user can still control the number of detections by changing the γ threshold similarly to the original detector (Figure 4e). Thus the value $\alpha = 0.2$ is used in all following experiments. The final threshold γ is the same in all experiments and is set empirically so that the detector produces similar number of detections as the original Hessian-Laplace detector.

Classifier length. Empirically we set the length of the classifier to $T = 20$ (number of weak classifiers). Longer classifiers slow down the evaluation (see Figure 4a) and do not bring significant improvement in performance.

Repeatability. The repeatability measure of the trained WaldBoost detector has been compared with the original Hessian-Laplace detector on standard image sequences with variations in scale and rotation, blur, affine deformation, light change and JPEG compression from [20]. The results are shown in Figure 5. The WaldBoost detector achieves



(a) (b) (c)

α	0.02	0.05	0.1	0.2	0.4
♯det	11472	9427	7106	4203	2624

(d)

threshold	500	1000
♯det	3549	2207

(e)

FIGURE 4. Selecting the false negative rate α. (a) The average evaluation speed for several values of α. All compared detectors are able to achieve similar number of correspondences (b) and repeatability score (c) – measured for $T = 20$ for all detectors on the BOAT sequence. (d) The number of detections of the WaldBoost emulator on the first image from the BOAT sequence as a function of the α parameter, (e) the number of detections of Hessian-Laplace as a function of the final threshold.

similar repeatability and number of correspondences as the original Hessian-Laplace detector.

Matching score. For the same sequences, the matching score of the trainer and the trainee is shown in Figure 6. The WaldBoost detector achieves slightly better matching score than the original algorithm.

Speed. The WaldBoost classifier evaluates on average 1.7 features per examined position and scale. Unsurprisingly, this is much less than any reported speed for face detection [30]. The evaluation times are compared in Table 1. The WaldBoost emulator is about nine times faster than the Hessian-Laplace detector with a rather careful design [22].

Classifier structure. The Hessian-Laplace detector finds blob-like structures. The structure of the trained WaldBoost emulation should reflect this property. As shown in Figure 7, the first selected weak classifier is of the centre-surround type and gives high responses to blob-like structures with high contrast between central part and its surrounding (the feature value is average intensity in the central part minus average intensity in the surrounding part).

Coverage. The output of the trained WaldBoost emulation of Hessian-Laplace is compared to the original algorithm in Figure 8a. As in the repeatability experiment two sets of detections are compared – the original detections and the WaldBoost emulator detections (with $\gamma = -\infty$). Since the comparison works on a single image, the ground truth transformation matrix is identity.

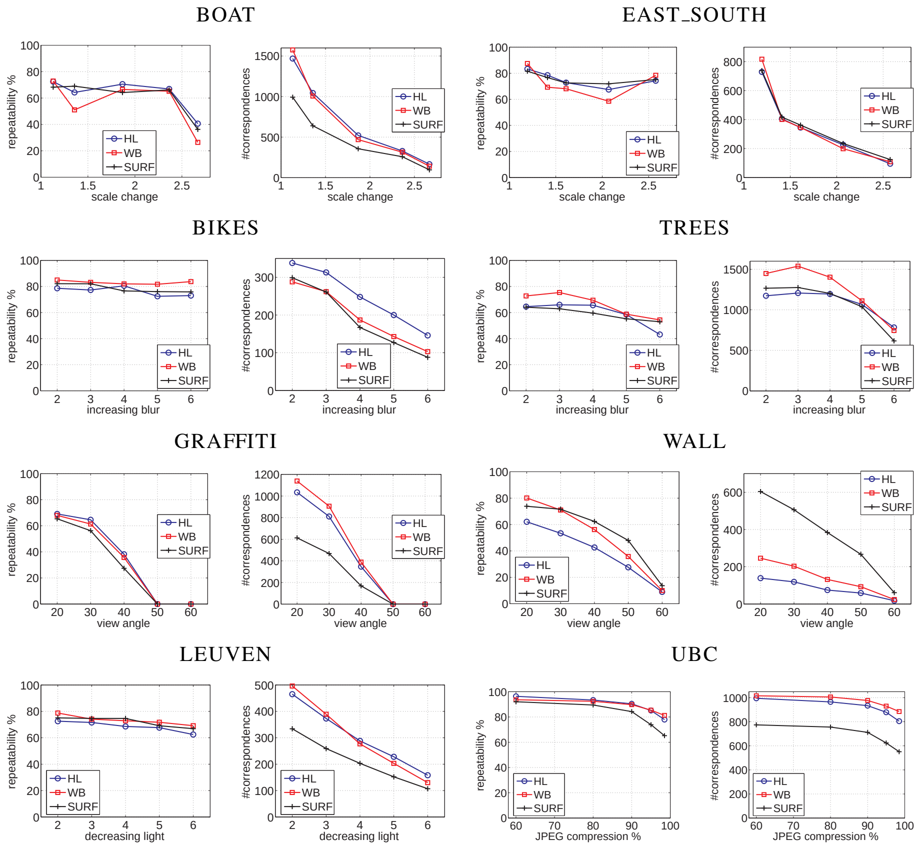


FIGURE 5. Repeatability comparison of the Hessian-Laplace detector, its WaldBoost emulation and the SURF detector on Mikolajczyk's dataset.

The white circles show the original detections with a correspondence found among the WaldBoost detections. The black circles show the original detections not found by WaldBoost. Note that most of the missed detections have a correct detection nearby, so the corresponding image structure is actually found. The percentage of repeated detections of the original algorithm is 80 %.

The WaldBoost detector may seem to miss consistently the large regions. Figure 8c shows manually found WaldBoost regions close to the original detections – the "tree blob" is in fact detected. The real problem is in the correspondence overlap computation. To compute the overlap of two detected points, Mikolajczyk [24] first normalises their scale to 30 pixels. This way, the problem of unnecessary large regions which would almost always have large overlaps is avoided. However, as shown in Figure 8d, this normalisation

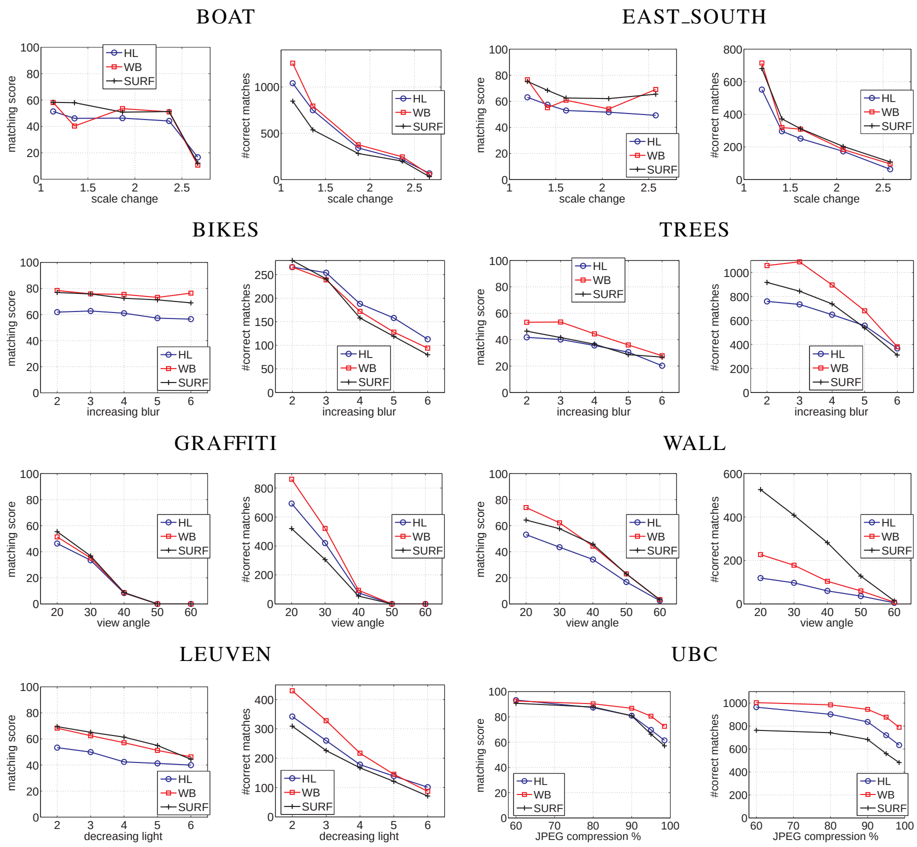


FIGURE 6. Matching score comparison of the Hessian-Laplace detector, its WaldBoost emulation and the SURF detector on Mikolajczyk's dataset.

returns small overlap when large regions are only slightly misplaced. This problem is general and *appears in all region detection papers which use the Mikolajczyk's repeatability measure*. To conclude, the real emulation accuracy is in fact higher than 80 %.

Rotational invariance. One of the properties of the emulated Hessian-Laplace detector which should be preserved is its rotational invariance. A learning approach can achieve rotational invariance with non-rotationally invariant features by introducing synthetically rotated positive samples into the training set. The results in Figure 10 (top row) show that the rotational invariance is preserved even without introducing synthetic training samples. This is probably a consequence of the large training pool which is available. Instead of introducing rotated samples synthetically, the statistics are covered by collecting huge number of samples.

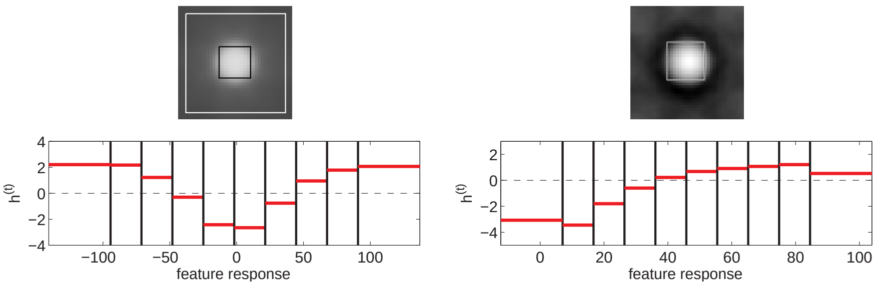


FIGURE 7. *Top row.* First centre-surround and variance feature found in WaldBoost Hessian-Laplace (left) and Kadir-Brady (right) emulated detectors. The background image is visualised as $E(|x_i - 127.5|)$ and $E(x_i)$ respectively, where $E()$ is the average operator and x_i is the i-th positive training example. *Bottom row.* Bin responses in the corresponding domain-partitioning weak classifiers (see Figure 1).

Scale invariance. Similarly, the detector invariance to scale changes has been tested. The emulated detector achieves similar scale invariance as the original algorithm as shown in Figure 10 (bottom row).

Comparison to SURF. The WaldBoost emulator has been compared with the SURF detector [2] which is a simplification of the Hessian-Laplace detector, manually designed for maximum speed. The SURF is commonly used as a good compromise between speed, accuracy and repeatability.

The comparison of the repeatability and the matching score of all three detectors is shown in Figure 5 and Figure 6. All the detectors has been set to produce similar number of detections on the first image of the EAST_SOUTH sequence. Neither of the fast detectors approximates the original detector perfectly. Yet, both could be said to achieve similar statistics as the original Hessian-Laplace detector, deviating slightly at different sequences.

The evaluation speeds of the detectors are compared in Table 1. The WaldBoost detector achieves similar evaluation speed as the manually tuned SURF detector. However, since most of the computational components are the same in both detectors, the average evaluation time $\bar{T}_{S^*} = 1.7$ for WaldBoost and $\bar{T}_{S^*} = 3$ for SURF suggests that further code optimisation of the WaldBoost detector could lead to even faster implementation.

An important difference between the SURF detector and the Hessian-Laplace WaldBoost emulator is that the first one is *a simplification* while the other is *an emulation*. The SURF produces different set of regions compared to the Hessian-Laplace detector. This could be verified by computing the coverage score as in Figure 8. For the SURF detector only 49.7 % coverage is reached compared to 80 % of the WaldBoost detector. The difference in detectors outputs is shown in Figure 9.

The WaldBoost emulator of the Hessian-Laplace detector is able to detect points with similar repeatability and slightly higher matching score while keeping the rotational

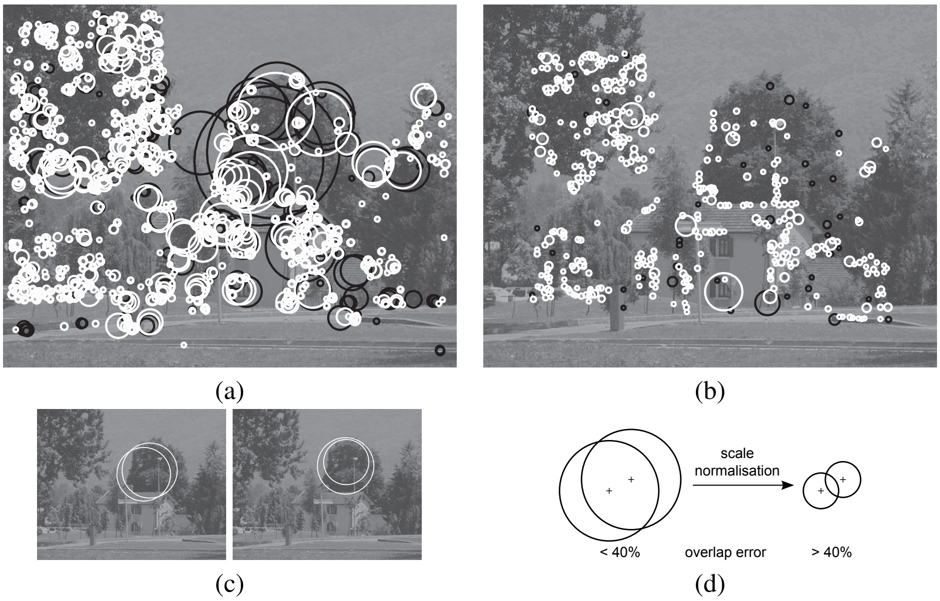
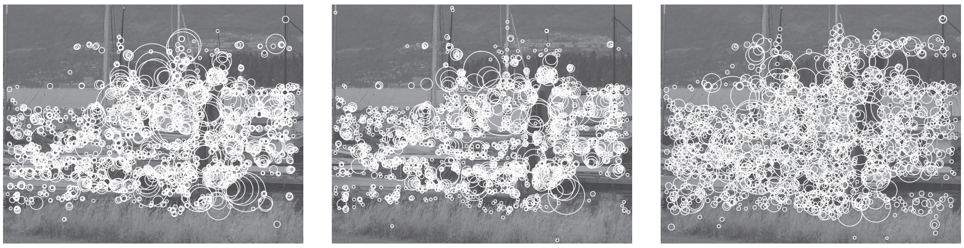


FIGURE 8. Comparison of the outputs of the original and WaldBoost-emulated (a) Hessian-Laplace and (b) Kadir-Brady saliency detectors. The white circles show repeated Hessian-Laplace detection. The black circles highlight the original detections not found by the WaldBoost detector. Note that for most of missed detections there is a nearby detection on the same image structure. The accuracy of the emulation is 80 % for Hessian-Laplace and 90 % for Kadir-Brady saliency detector. Note that the publicly available Kadir-Brady algorithm does not detect points close to image edges. (c) Missed Hessian-Laplace detections (left) and manually found corresponding WaldBoost detections (right). (d) They are not found as correspondences, because Mikolajczyk's overlap function prefers smaller detections (see the discussion in the text).

and scale invariance of the original detector. Moreover, the WaldBoost emulator was able to increase nine times the speed of detection compared to the original detector. When compared to the manually tuned SURF detector, similar repeatability, matching score and evaluation speed characteristics are reached. However the WaldBoost detector emulates the Hessian-Laplace detector significantly more closely.

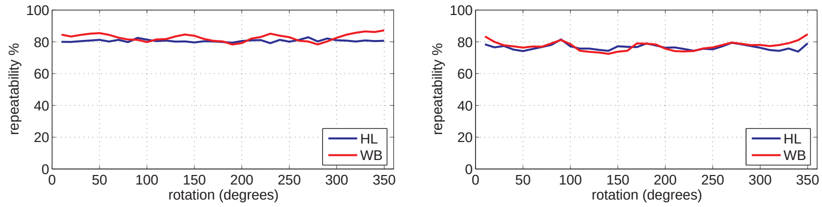
5.2. Fast saliency detector

The emulation of the Kadir-Brady saliency detector [14] uses the same image pool for training as the WaldBoost Hessian-Laplace emulator. The saliency threshold of the original detector was set empirically to 2 to collect a sample pool of a reasonable size. Higher value of threshold also helps to limit the positive examples only to those with higher

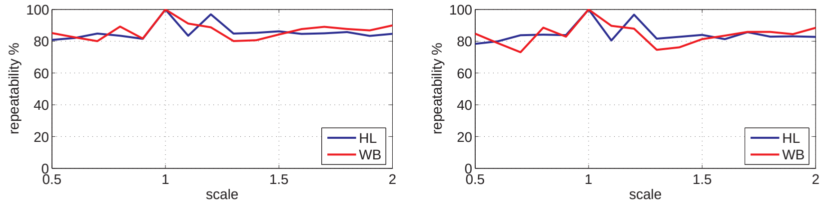


(a) (b) (c)

FIGURE 9. Comparison of Hessian-Laplace (a), its WaldBoost emulator (b) and SURF detector (c) outputs on the first image from the BOAT sequence. WaldBoost returns similar distribution of points as the emulated Hessian-Laplace. The SURF points are distributed differently.



(a) (b)



(c) (d)

FIGURE 10. Rotation and scale invariance of the WaldBoost Hessian-Laplace emulator. *Top row:* Repeatability on *rotated* first images from (a) BOAT, and (b) EAST_SOUTH sequences for the Hessian-Laplace detector (HL) and its WaldBoost emulator (WB). *Bottom row:* Repeatability on *scaled* first images from (c) BOAT, and (d) EAST_SOUTH sequences.

saliency. As opposed to the Hessian-Laplace emulation, where rather low threshold was chosen, it is meaningful to use only the top most salient features from the Kadir-Brady detector since its response corresponds to the importance of the feature.

	Hessian-Laplace	**Kadir-Brady**
original	0.9s	1m 48s
SURF	0.09s	—
speed-up	10×	—
T_{s^*}	3	—
WaldBoost	0.10s	0.76s
speed-up	9×	142×
T_{s^*}	1.7	2.2

TABLE 1. Speed comparison on the first image (850×680) from the BOAT sequence. The speed-up on another images is similar.

The Haar-like feature set was extended by the "variance" feature described in Section 4. The training was run for $T = 20$ (training steps) with $\alpha = 0.2$ and $\beta = 0$ as in the Hessian-Laplace experiment.

Publicly available version of Kadir-Brady detector has several drawbacks which need to be considered in the experimental evaluation. Due to relatively wide search for local maximum in the scale space, detections near the image border are not detected. This results in a strip around image border where no detections are returned (see Figure 8b). Also the scale range of detections is limited. In all following experiments, WaldBoost emulator detections are filtered by the same restrictions for the comparison reasons. However, the WaldBoost emulator of the Kadir-Brady detector *does not have these restrictions* inherently.

Repeatability and matching score. The same experiments as for the Hessian-Laplace detector have been performed. The repeatability and the matching score of the Kadir-Brady detector and its WaldBoost emulation on BOAT and EAST_SOUTH sequences are shown in Figure 11. Similar performance to the teacher is reached for similar number of correspondences and correct matches on both sequences.

Speed. The main advantage of the emulated saliency detector is its speed. The classifier evaluates on average 2.2 features per examined location and scale. Table 1 shows that the emulated detector is about 142× faster than the original detector.

Classifier structure. Our early experiments showed that the Haar-like features are not suitable to emulate the entropy-based saliency detector. With the variance features, the training was able to converge to a reasonable classifier. In fact, the variance feature is chosen for the first weak classifier in the WaldBoost ensemble (see Figure 7). The bin responses of the weak classifier show that higher variations are preferred.

Coverage. The outputs of the WaldBoost saliency detector and the original algorithm are compared in Figure 8b. The coverage of original detections is 90 %.

Rotational and scale invariance. Invariance to rotation and scale changes of the WaldBoost emulator and the Kadir-Brady detector are compared in Figure 13. Due to

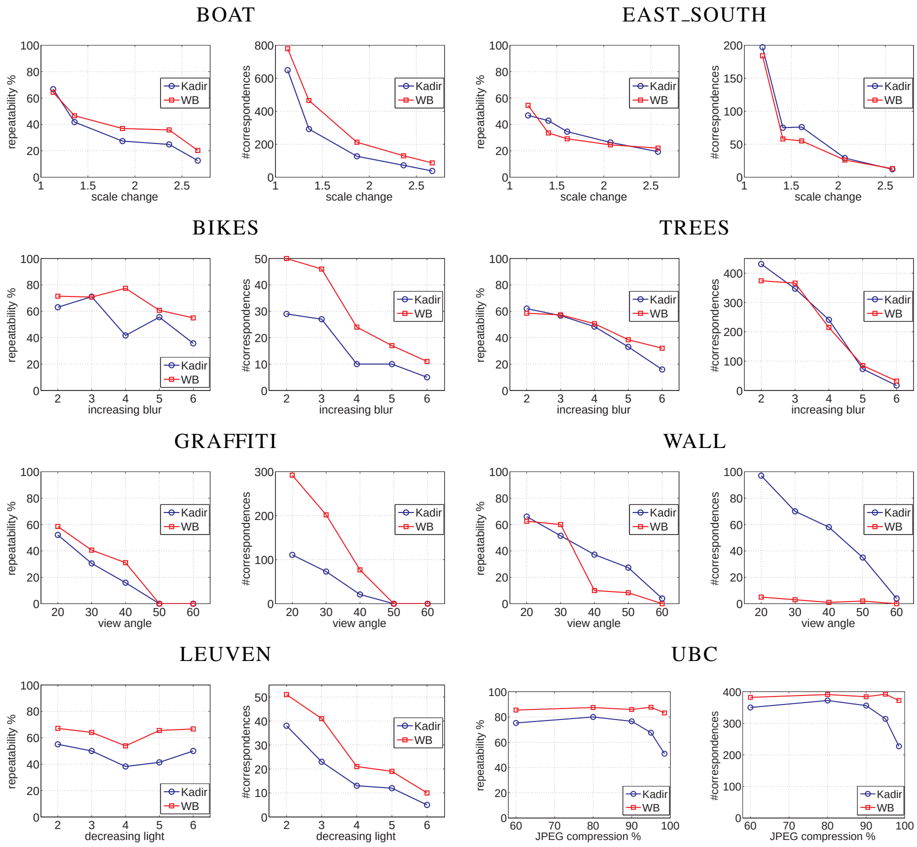


FIGURE 11. Repeatability comparison of the Kadir-Brady detector and its WaldBoost emulation on Mikolajczyk's dataset.

very different approaches in computing the detectors responses (Haar-like features vs. entropy), the WaldBoost emulator is not able to reach perfect rotation invariance on images rotated by 90 degrees but is able to keep similar rotational invariance otherwise. Moreover, the feature-based approach of the WaldBoost emulator results in slightly better scale invariance of the detector. This can be probably explained by the instability of the entropy based Kadir-Brady detector especially at small scales where the probabilities are difficult to estimate. It is shown also in [11] that their difference-of-Gaussians detector is more robust to a range of transformations than the Kadir-Brady detector.

To conclude, the WaldBoost training is able to emulate Kadir-Brady detector generally with similar repeatability, matching score and robustness to rotation changes, while improving slightly its scale invariance. But, most importantly, the decision times of the

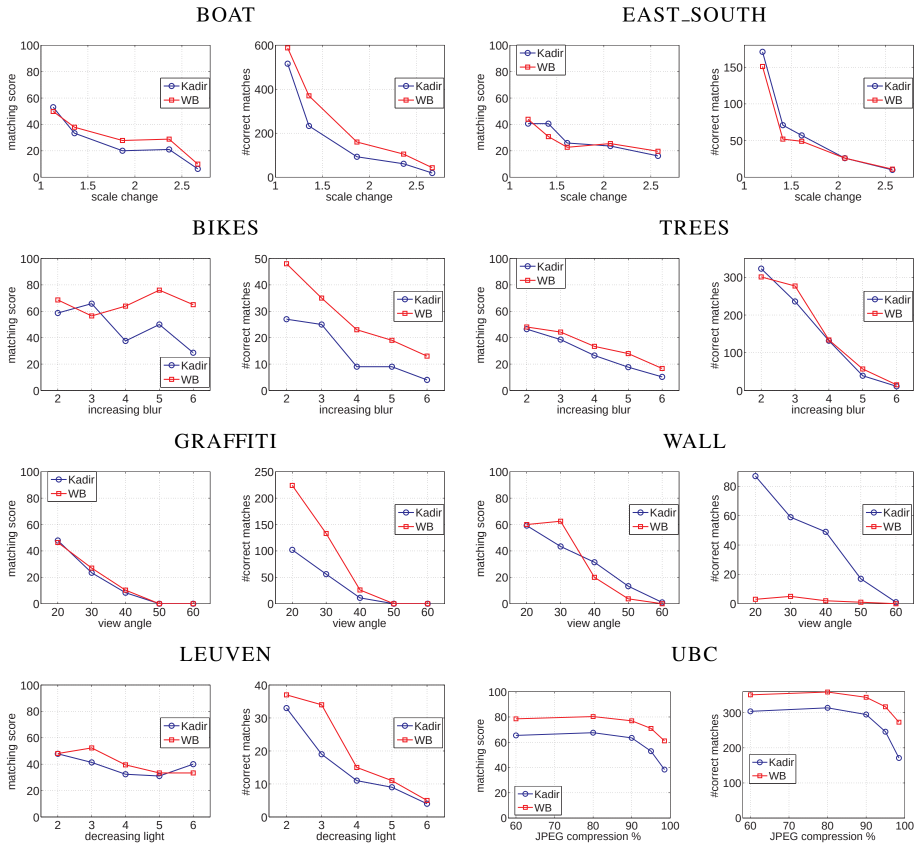


FIGURE 12. Matching score comparison of the Kadir-Brady detector and its WaldBoost emulation on Mikolajczyk's dataset.

emulated detector are about 142 times lower than that of the original algorithm. That opens new possibilities for using the Kadir-Brady detector in time sensitive applications.

6. Conclusions and future work

We have presented a method for learning sequential two-class classifiers with decision quality and evaluation time trade-off. The method – a learning algorithm called WaldBoost – enlarges SPRT's applicability to problems with dependent measurements and removes the limitations of SPRT to a priori ordered measurements and known joint probability densities. Asymptotic properties of the AdaBoost learning algorithm are exploited for selection and ordering of relevant measurements that are subsequently processed by SPRT.

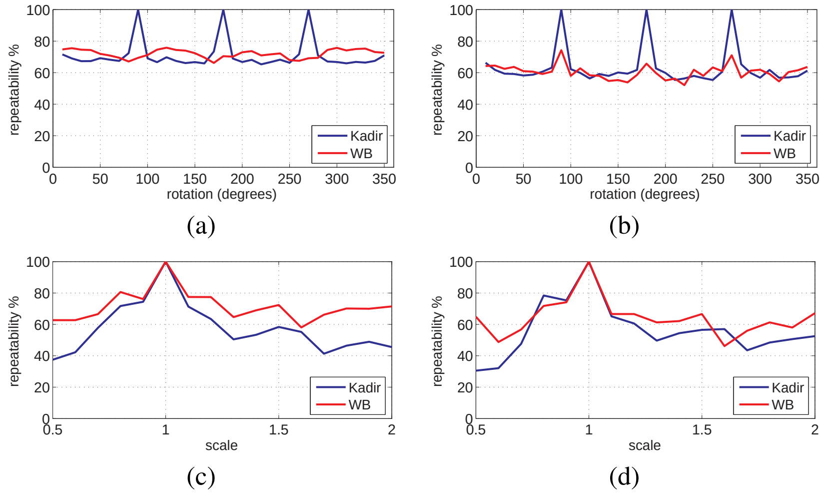


FIGURE 13. Rotation and scale invariance of the WaldBoost Kadir-Brady emulator. *Top row:* Repeatability on *rotated* first images from (a) BOAT, and (b) EAST_SOUTH sequences for the Kadir-Brady detector (Kadir) and its WaldBoost emulator (WB). *Bottom row:* Repeatability on *scaled* first images from (c) BOAT, and (d) EAST_SOUTH sequences.

One possible application of the WaldBoost algorithm, a framework for speeding up existing binary decision processes by learning their WaldBoost emulators, was presented. Two interest point detectors, the Hessian-Laplace and the Kadir-Brady saliency detector, served as examples of emulated algorithms. The experiments show similar repeatability and matching scores of the original and emulating algorithms. For both, the Hessian-Laplace and the Kadir-Brady detectors, the WaldBoost emulation improved significantly the speed. The emulator was nine times faster for the Hessian-Laplace detector and about 142 times faster for the Kadir-Brady detector. In the case of the Kadir-Brady detector this speed-up opens new possibilities for using the detector in time sensitive applications. For the Hessian-Laplace detector, the achieved speed is similar to SURF, a commonly used Hessian-like fast detector; the WaldBoost emulator approximates the output of the Hessian-Laplace detector more precisely.

The proposed emulation approach is general and can be applied to other algorithms as well. For future research, an interesting extension of the methodology would be to train an emulator which not only guarantees output similar to an existing algorithm but which also possesses some additional quality like insensitivity to certain acquisition conditions (e.g. motion blur) or maximum performance in a particular environment or task.

Returning to the WaldBoost framework, future research will include generalization to multiclass decision problems as well as consideration of measurements with variable costs.

References

[1] Baker S, Nayar SK (1996) Algorithms for pattern rejection. In: ICPR, IEEE Computer Society, Washington, DC, USA, p 869
[2] Bay H, Ess A, Tuytelaars T, Gool LV (2008) Speeded-up robust features (SURF). Computer Vision and Image Understanding 110(3):346–359
[3] Bourdev L, Brandt J (2005) Robust object detection via soft cascade. In: CVPR, IEEE Computer Society, Washington, DC, USA, pp 236–243
[4] Brubaker SC, Wu J, Sun J, Mullin MD, Rehg JM (2008) On the design of cascades of boosted ensembles for face detection. IJCV 77(1-3):65–86, DOI http://dx.doi.org/10.1007/s11263-007-0060-1
[5] Dollár P, Tu Z, Belongie S (2006) Supervised learning of edges and object boundaries. In: CVPR, vol 2, pp 1964–1971
[6] Fergus R, Perona P, Zisserman A (2005) A sparse object category model for efficient learning and exhaustive recognition. In: CVPR, vol 1, pp 380–387
[7] Friedman J, Hastie T, Tibshirani R (1998) Additive logistic regression: a statistical view of boosting. Tech. rep., Department of Statistics, Sequoia Hall, Stanford Univerity
[8] Froba B, Ernst A (2004) Face detection with the modified census transform. In: AFGR, pp 91–96
[9] Grabner H, Šochman J, Bischof H, Matas J (2008) Training sequential on-line boosting classifier for visual tracking. In: ICPR
[10] Grabner M, Grabner H, Bischof H (2006) Fast approximated SIFT. In: ACCV, pp I:918–927
[11] Hare JS, Lewis PH (2004) Salient regions for query by image content. In: Image and Video Retrieval: Third International Conference, Dublin, Ireland, pp 317–325
[12] Huang C, Ai H, Lao S, Li Y (2007) High-performance rotation invariant multiview face detection. IEEE Trans Pattern Anal Mach Intell 29(4):671–686, DOI http://dx.doi.org/10.1109/TPAMI.2007.1011
[13] Jones MJ, Rehg JM (2002) Statistical color models with application to skin detection. Int J Comput Vision 46(1):81–96
[14] Kadir T, Brady M (2001) Saliency, scale and image description. IJCV 45(2):83–105
[15] Kálal Z, Matas J, Mikolajczyk K (2008) Weighted sampling for large-scale boosting. In: BMVC
[16] Lepetit V, Lagger P, Fua P (2005) Randomized trees for real-time keypoint recognition. In: CVPR, vol II, pp 775–781
[17] Lienhart R, Maydt J (2002) An extended set of haar-like features for rapid object detection. In: ICIP, vol 1, pp 900–903
[18] Martin D, Fowlkes C, Malik J (2004) Learning to detect natural image boundaries using local brightness, color, and texture cues. PAMI 26(5):530–549
[19] Matas J, Šochman J (2007) Wald's sequential analysis for time-constrained vision problems. In: Hutchinson S (ed) IEEE International Conference on Robotics and

Automation, Workshops and Tutorials, IEEE, CDROM, invited paper
[20] Mikolajczyk K (2002) Detection of local features invariant to affines transformations. PhD thesis, INPG, Grenoble
[21] Mikolajczyk K (2008) http://www.robots.ox.ac.uk/˜vgg/research/affine
[22] Mikolajczyk K (2008) Personal communication
[23] Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. IJCV 60(1):63–86
[24] Mikolajczyk K, Tuytelaars T, Schmid C, Zisserman A, Matas J, Schaffalitzky F, Kadir T, Van Gool L (2005) A comparison of affine region detectors. IJCV 65(1/2):43–72
[25] Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 24(7):971–987
[26] Rivest RL (1987) Learning decision lists. In: Machine Learning, pp 229–246
[27] Rosten E, Drummond T (2006) Machine learning for high-speed corner detection. In: European Conference on Computer Vision, vol 1, pp 430–443
[28] Schapire RE, Singer Y (1999) Improved boosting algorithms using confidence-rated predictions. Machine Learning pp 37(3): 297–336
[29] Siegmund D (1985) Sequential Analysis. Test and Confidence Intervals. Springer Series in Statistics, Springer-Verlag, New York, NY
[30] Šochman J, Matas J (2005) WaldBoost - learning for time constrained sequential detection. In: CVPR, Los Alamitos, USA, vol 2, pp 150–157
[31] Šochman J, Matas J (2007) Learning a fast emulator of a binary decision process. In: ACCV
[32] Sung KK, Poggio T (1998) Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1):39–51
[33] Viola P, Jones M (2001) Robust real time object detection. In: SCTV, Vancouver, Canada
[34] Wald A (1947) Sequential analysis. Dover, New York
[35] Xiao R, Zhu L, Zhang H (2003) Boosting chain learning for object detection. In: ICCV, pp 709–715
[36] Zhu L, Chen Y, Yuille AL (2006) Unsupervised learning of a probabilistic grammar for object detection and parsing. In: NIPS, pp 1617–1624

Curriculum Vitae

Doc. Dr. Ing. Jiří Matas
Date of birth: August 6, 1964

Education and Academic Qualifications

2005	Assoc. Professor (doc.), CTU Prague, in Cybernetics.
1995	PhD degree, University of Surrey, UK. Advisor: Prof. J. Kittler.
1987	MSc degree (with honours) in electrical engineering, CTU Prague.

Employment

2006-2009	Assoc. Prof., Center for Machine Perception, CTU Prague, CZ.
2007	Visiting Professor, EPFL Lausanne, Switzerland.
2005-2006	Visiting Researcher, CVSSP group, University of Surrey, UK.
1997-2005	Senior Researcher, Center for Machine Perception, Prague, CZ.
1997-2001	I concurrently held two part-time positions, at the CVSSP, U.of Surrey, UK; and Center for Machine Perception, CTU Prague, CZ.
1991-1997	Research fellow, CVSSP, U. of Surrey, UK.
1990	Visiting fellow, Centre for Vision, Speech and Signal Processing (CVSSP), U. of Surrey, UK.
1987-1990	Department of Control, CTU Prague, CZ.
1986	Statistical Department of EDP Coimbra, Portugal. Two-month employment organised by the International Association For The Exchange Of Students For Technical Experience.

Awards

2009	K. Zimmermann's PhD thesis "Fast Learnable Methods for Object Tracking" supervised by J. Matas was awarded the A. Svoboda prize for the "Best PhD dissertation in Czech Republic in the fields of cybernetics and informatics in 2008".
2007	The paper "J. Sochman, J. Matas: Learning A Fast Emulator of Binary Decision Process" was awarded the best paper prize at the Asian Conference on Computer Vision
2005	The Center for Machine Perception team that I was a member of finished second in the ICCV 2005 Contest
2005	The paper "J. Matas, S. Obdrzalek: Sub-linear Indexing for Large Scale Object Recognition" was awarded the best paper prize at the British Machine Vision Conference, Oxford, UK.
2004	"The Best Scientific Result" prize of the Czech Technical University Prague.
2002	The paper "J. Matas et al.: Robust wide baseline stereo from maximally stable extremal regions" was awarded the best paper prize at the British Machine Vision Conference, Cardiff, UK.
1987	MSc. thesis awarded the Chancellor's prize.

Grants, Research Projects
Principal investigator of the "Algorithms for Face recognition" project (2001-2003) funded by the Czech Grant Agency (the project was evaluated as "excellent"). Principal investigator of the Czech Grant Agency project "Methods for Visual Recognition of Large Collections of Non-rigid Objects" (till 2011).
Participated in a number of EU funded projects as a researcher and later as a work package leader, working on following problems: recognition in active vision systems (1991-1995, VAP "Vision as Process" project), face recognition and lip tracking (1995-1997, M2VTS "Multimodal Verification for Teleservices and Security Applications), for face recognition, lip tracking and speaker verification, biometric identity verification (1997-1999, BANCA "Biometric access control for networked and e-commerce applications"), visual recognition (2001-2004, ActIpret "Activity Interpretation") and COSPAL (2004-2007, "COgnitiveSystems using Perception-Action Learning") and DIPLECS (2007-2010, "Dynamic Interactive Perception-action LEarning in Cognitive Systems"). The principal investigator of FP7 EU project MASH (2010-2012, "Massive sets of heuristics").

Industrial Applications. Consultancies.
Co-founder of a university spin-off company Eyedea Recognition (established in 2006).
Lead many industry-sponsored projects and consulted for a number of companies, selected examples:

Hitachi, Japan	2003-2009	Project leader. Face analysis.
Toyota, Japan	2003-2010	Project leader. Object recognition in traffic applications.
Samsung, South Korea	2001-2004	Project leader. Face recognition.
VUL Prague, CZ	2001-2002	Project leader. Computer vision techniques for processing of video data acquired by sensors on the airplane.
Boeing, USA	1999-2000	Project leader. Development of software for airplane recognition.
Racal, UK	1996	Consultant in a project on the use of colour for license plate recognition.
Zbrojovka Brno, Czechoslovakia	1987-1988	Head developer. Image processing software for a machine vision system.
VUJE Trnava, Czechoslovakia	1988	Project leader. Software development for automatic positioning of a defectoscopic ultrasonic testing device in a nuclear power plant.
TESLA Roznov, Czechoslovakia	1989	Image processing specialist. Development of an image processing system for control of the silicon rafination process. of integrated circuits in Czechoslovakia.

Other Professional Activities.

- Associate Editor-in-Chief of IEEE Transaction on Pattern Analysis and Machine Intelligence (highest impact factor of all IEEE Transaction and all journals in ISI in both Electrical Engineering and Artificial Intelligence categories).
- Editorial board member of International Journal of Computer Vision and Pattern Recognition (number 2 and 3 journals in the field).

- Evaluator for: the EU FP6 IST call (2003), the EU FP7 IST call (2007), EU Marie-Curie fellowship applications (2007, 2008, 2009). EU FET (Future and Emerging Technologies) projects (2009), the Grant Agency of Hong Kong (2008 and 2009), Swiss and Swedish grant agencies.
- Evaluation panel member for the Czech Grant Agency (since 2009).
- (2003) Representative of the Czech Republic in the standardisation body ISO-IEC JTC 1SC 29WG 11(MPEG). Participated in the proposal of a face descriptor that was included in the MPEG standard.
- Programme committee of a number of major international conferences in the area of computer vision, image retrieval and pattern recognition: Int. Conf. on Comp. Vision, Comp. Vision and Pattern Recognition, Int. Conference on Pattern Recognition, Neural Information Processing Systems, European Conf. on Comp. Vision, International Conf. on Face and Gesture Recognition, Audio- and Video-based Biometric Person Authentication, International Conf. on Image and Video Retrieval, and other.
- Programme co-chair of the European Conference on Computer Vision in 2004 and the Computer Vision and Pattern Recognition Conference in 2007.
- Chairman of Technical Committee 14 "Signal Analysis for Machine Intelligence" of the International Association for Pattern Recognition(2002 - 2004).

Teaching.

2005	Image Processing and Vision, MSc module, University of Surrey, UK
1997-2009	Pattern Recognition, MSc Course, CTU Prague, Czech R.
2000-2008	Digital Image Processing, MSc Course, CTU Prague, Czech R.
2002-2009	Advanced Pattern Recognition, PhD Course, CTU Prague, Czech R.

PhD supervision.
From 1996 till 2003 co-supervised 4 PhD students at the University of Surrey. All four successfully defended their thesis. At CMP Prague 4 PhD students supervised by J. Matas graduated. Currently supervising or co-supervising 3 PhD students.

Publications and Citations
$>$ 150 papers in refereed journals and conferences.
$>$ 2800 citations in the ISI WoS.
h-index: 19 (ISI WoS).

Selected publications:
(most publications are available on-line at `http://cmp.felk.cvut.cz/~matas`).

- J. Sochman and J. Matas. Learning Fast Emulators of Binary Decision Processes. International Journal of Computer Vision. 83(2), 2009, pp. 149-163.
- K. Zimmermann, J. Matas, and T. Svoboda. Tracking by an Optimal Sequence of Linear Predictors. IEEE Transactions on Pattern Analysis and Machine Intelligence. 31(4), 2009, pp. 677-692.

- S. Obdrzalek and J. Matas. Object Recognition using Local Affine Frames on Maximally Stable Extremal Regions. Invited chapter in J. Ponce et al., editors: Toward Category-Level Object Recognition, Springer. 2006, pp. 85-108.
- Chum, J. Matas. Optimal Randomized RANSAC. IEEE Transactions on Pattern Analysis and Machine Intelligence. 30(8), 2008, pp. 1472-1482.
- J. Matas, O. Chum, U. Martin, and T Pajdla. Robust wide baseline stereo from maximally stable extremal regions. Image and Vision Computing. 22(10), 2004, pp. 761-767.
- J. Matas, C. Galambos, and J. Kittler. Robust detection of lines using progressive probabilistic Hough Transform. Computer Vision and Image Understanding. 78(1), 2000, pp. 119-137.
- J. Matas, K. Jonsson, and J. Kittler. Fast face localisation and verification. Image and Vision Computing. 17(8), 1999, pp. 578-581.
- J. Kittler, M. Hatef, R.P.W Duin, and J. Matas. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(3), 1998, pp. 226-239.