Improving object detection with boosted histograms

Ivan Laptev

INRIA Rennes - Bretagne Atlantique
Campus universitaire de Beaulieu
35042 Rennes Cedex, France
ivan.laptev@inria.fr

Abstract

We address the problem of visual object class recognition and localization in natural images. Building upon recent progress in the field we show how histogram-based image descriptors can be combined with a boosting classifier to provide a state of the art object detector. Among the improvements we introduce a weak learner for multi-valued histogram features and show how to overcome problems of limited training sets. We also analyze different choices of image features and address computational aspects of the method. Validation of the method on recent benchmarks for object recognition shows its superior performance. In particular, using a single set of parameters our approach outperforms all the methods reported in VOC05 Challenge for 7 out of 8 detection tasks and four object classes while providing close to real-time performance.

Key words: object recognition, machine learning, histogram image features

1 Introduction

Among the vast variety of existing approaches to object recognition there is a remarkable success of methods using histogram-based image descriptors. An influential work by Swain and Ballard [26] proposed colour histograms as a simple and efficient image descriptor for object recognition. The idea was further developed by Schiele and Crowley [23] who recognised objects using histograms of local filter responses. Histograms of Textons were proposed by Leung and Malik [14] as well as by Varma and Zisserman [27] for texture recognition. Schneiderman and Kanade [24] computed histograms of wavelet coefficients over localised object parts and were among the first to address

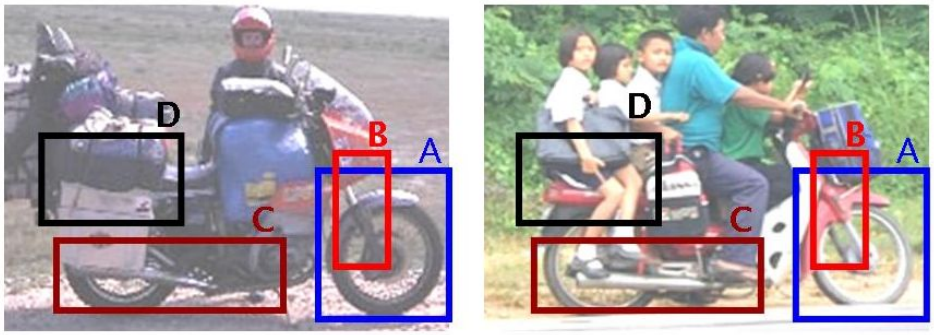

Fig. 1. Rectangles on the left and right image are examples of possible regions for histogram features. Stable appearance in A,B and C on both images makes corresponding features to be good candidates for a motorbike classifier. On the contrary, regions D are unlikely to contribute for the classification due to the large variation in appearance.

object class detection in images of natural scenes. In a similar spirit the well-known SIFT descriptors [18] and Shape Context [1] as well as more recent HOG descriptor [2] and Spatial Pyramid representations [13] make an effective use of position-dependent histograms to describe local and global image content.

Histograms represent distributions of spatially unordered image measurements in a region and provide relative invariance to several variations of object appearance. The invariance and the descriptive power of histograms, however, crucially depend on (i) the type of local image measurements and (ii) the image regions used to accumulate histograms. Regarding the type of measurements, different alternatives have been proposed and investigated that may have better performance depending on the task [26,23]. As a general purpose image descriptor, the choice of Histograms of Oriented Gradients (HOG) is well supported by successful applications of SIFT descriptor [18,21] and other related methods [2].

Besides the question *what* to measure, the question *where* to measure obviously has a large impact on recognition performance. Global histograms [26,23] have recently achieved impressive performance for scene categorization [13,31]. Object recognition and localization, however, is currently better addressed by local methods [24,18,2] computing histograms over local image regions. As illustrated in Figure 1, different regions of an object may have different descriptive power and, hence, different impact on the learning and recognition. In the previous work histogram regions were often selected either a-priori using fixed grids [24,2] or by applying region detectors of different kinds [18,3,19]. None of these two alternatives, however, guarantees an optimal choice of histogram regions for subsequent recognition. An arguably more attractive approach proposed by Levi and Weiss [15] and confirmed in [12,32] consists of learning class-specific histogram regions from the training data. We follow this

approach and note its conceptual similarity to other methods making attempt to discover discriminative object parts for visual recognition [5].

In this work, similar to [15], we select the position and the shape of histogram features to minimise the training error for a given recognition task. During training, we consider an exhaustive set of rectangular regions in the normalised object window and compute histogram descriptors for each of them efficiently using integral histograms [22]. We then apply AdaBoost [8,29] to select histogram features and to learn an object classifier. As a part of our contribution to object learning, we adapt the boosting framework to vector-valued histogram features and design a weak learner based on Weighted Fischer Linear Discriminant (WFLD). We in addition deploy position-dependent histogram features and artificially enlarge the size of the training set by adding spatial noise to the annotation. These extensions demonstrate a substantial improvement with respect to [15].

To validate the proposed method, we test it on the task of object detection in natural images and evaluate the performance on PASCAL Visual Object Category datasets VOC 2005 and VOC 2006 [7,6]. Using a single set of parameters we demonstrate our approach to outperform all methods reported in the competition [7] for 7 out of 8 detection tasks and four object classes. Among the advantages of the method we emphasise (i) its ability to learn from a small number of samples, (ii) stable performance for different object classes, and (iii) close to real-time performance.

We further investigate the framework by comparing performance of alternative histogram features and feature selection mechanisms. Evaluation on several object classes confirms the high performance of HOG descriptors, however, the best performance is demonstrated by the combination of HOG features with other histogram descriptors in terms of second-order image derivatives and color. Given the popularity of interest point features in recognition methods, we also compare regions selected by our method with Harris-Affine regions [20]. Notably, we find Harris-Affine regions to perform no better than random regions in our framework tested on three different object classes. We finally investigate computational aspects of the method and evaluate its precision-speed trade-off.

The rest of the paper is organised as follows. In Section 2 we recall AdaBoost algorithm and develop a weak learner for vector-valued features. Section 3 defines histogram features and integrates them with the boosting framework. In Section 4 we evaluate and compare the method on the task of object detection. Sections 5 and 6 investigate alternative image features and computational aspects of the method respectively. Section 7 concludes the paper.

2 AdaBoost learning

AdaBoost [8] is a popular machine learning method combining properties of an efficient classifier and feature selection. The discrete version of AdaBoost defines a strong binary classifier H

$$H(z) = \mathrm{sgn}(\sum_{t=1}^{T} \alpha_t h_t(z))$$

using a weighted combination of T weak learners h_t with weights α_t. At each new round t, AdaBoost selects a new hypothesis h_t that best classifies training samples with high classification error in the previous rounds. Each weak learner

$$h(z) = \begin{cases} 1 \text{ if } \quad g(f(z)) > threshold \\ -1 \text{ otherwise} \end{cases} \tag{1}$$

may explore any feature f of the data z. In the context of visual object recognition it is attractive to define f in terms of local image properties over image regions r and then use AdaBoost for selecting features maximising the classification performance. This idea was first explored by Viola and Jones [29] who used AdaBoost to train an efficient face detector by selecting a discriminative set of local Haar features. Here similar to [15], we will define f in terms of histograms computed for rectangular image regions on the object.

2.1 Weak learner

The performance of AdaBoost crucially depends on the choice of weak learners h. While effective weak learners will increase the performance of the final classifier H, the potentially large number of features f prohibits the use of complex classifiers such as Support Vector Machines or Neural Networks. For one-dimensional features $f \in \mathbb{R}$ such as Haar features in [29], an efficient classifier for n training samples can be found by selecting an optimal decision threshold in (1) in $O(n \log n)$ time. For vector-valued features $f \in \mathbb{R}^m$ such as histograms, however, finding an optimal linear discriminant would require unreasonably long $O(\binom{n}{m})$ time.

One approach to deal with multi-dimensional features used in [15] is to project f onto a *pre-defined* set of 1-dimensional manifolds using a fixed set of functions $g_j \colon \mathbb{R}^m \to \mathbb{R}$. A weak learner can then be constructed for each combination of basis functions g_j and features f_i. Although efficient, such an approach can be suboptimal if a chosen set of functions g_j is not well suited for a given classification problem. As an example of inefficient AdaBoost classifier consider the problem of separating two diagonal distributions of points in $\mathbb{R}^2$ illustrated in Figure 2(left). Using axis-parallel linear basis functions $g_1(f) = (1 \;\; 0)f$ and

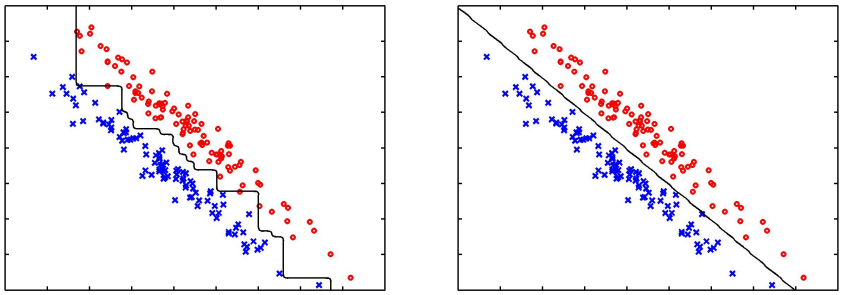

Fig. 2. Classification of two diagonal distributions using (left): AdaBoost with weak learners in terms of axis-parallel linear classifiers; (right): Fisher linear discriminant.

$g_2(f) = (0 \;\; 1)f$, the resulting AdaBoost classifier has poor generalisation and requires $T \approx 50$ weak hypotheses for separating $n = 200$ training samples.

An alternative and efficient choice for a multi-dimensional classifier is Fisher Linear Discriminant (FLD) [4]. FLD has been used as a weak learner in the context of AdaBoost in [30]. FLD guarantees optimal classification of normally distributed samples of two classes using a linear projection function

$$g = w^\top f \quad \text{with} \quad w = (S^{(1)} + S^{(2)})^{-1}(\mu^{(1)} - \mu^{(2)}) \tag{2}$$

defined by the class means $\mu^{(1)}$, $\mu^{(2)}$ and the class covariance matrices $S^{(1)}$, $S^{(2)}$. Illustration of FLD classification in Figure 2(right) clearly indicates its advantage in this example compared to the classifier in Figure 2(left). A particular advantage of using FLD as a weak learner is the possibility of re-formulating FLD to minimise a *weighted* classification error as required by AdaBoost. Given the weights d_i corresponding to samples z_i, the Weighted Fischer Linear Discriminant (WFLD) can be obtained using a function g in (2) with the means μ and covariance matrices S substituted by the weighted means μ_d and the weighted covariance matrices S_d defined as

$$\mu_d = \frac{1}{n \sum d_i} \sum_i^n d_i f(z_i), \quad S_d = \frac{1}{(n-1) \sum d_i^2} \sum_i^n d_i^2 (f(z_i) - \mu_d)(f(z_i) - \mu_d)^\top. \tag{3}$$

Using WFLD as an AdaBoost weak learner eliminates the need of re-sampling training data required by classifiers that do not make use of sample weights.

In practice, the distribution of image features $f(x_i)$ will mostly be non-Gaussian and multi-modal. Given a large set of features f, however, we can assume that the distribution of samples at least for some features will be close to Gaussians yielding the good performance of the resulting WFLD classifier. Experimental validation of this assumption and the advantage of WFLD will be demonstrated in Section 4 on real classification problems. In this work we use WFLD to find 1-dimensional projections of histogram features according to (2) and then determine an optimal classification threshold as in [29].

3 Image features

During training we assume a rough alignment of object samples within a rectangular window (see Figure 5). Under this assumption we rely on the correspondence of object parts and learn the appearance of parts from corresponding image regions. To avoid a heuristic selection of such regions, we initially consider an exhaustive set of rectangular sub-windows r on the object for AdaBoost learning as illustrated in Figure 3(Left).

3.1 Histogram features

We represent each feature by a histogram of local image measurements within a region r. Following previous work [18,2], we initially adopt Histograms of Oriented Gradients (HOG) features and consider histograms of alternative image measurements such as color and second order image derivatives later in Section 5. To construct HOG features, we compute orientation γ of local image gradient at each point $(x, y) \in r$

$$\gamma(x, y) = \arctan \frac{L_x(x, y)}{L_y(x, y)}, \quad L_\xi = I * \frac{\partial}{\partial_\xi} \left(\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \right) \Big|_{\xi=x|y} \tag{4}$$

using Gaussian derivatives L_x, L_y [16] of image I computed for scale parameter σ. We discretize γ into $m = 4$ equal orientation bins and increment histograms by the values of the gradient magnitude $||(L_x, L_y)||_2$. The histograms are normalized to the l_1 unit norm.

To preserve rough location of image measurements within a region, we subdivide regions into parts as illustrated in Figure 3(Right) and compute histograms separately for each part. Four types of image features $f_{k,r}(I)$ with spatial grids $k = \{1\text{x}1, 1\text{x}2, 2\text{x}1, 2\text{x}2\}$ are then computed for each region r by concatenating part-histograms into feature vectors of dimensions $m, 2m, 2m$ and $4m$ respectively. We use integral histograms [15,22] for efficient computation of histogram features.

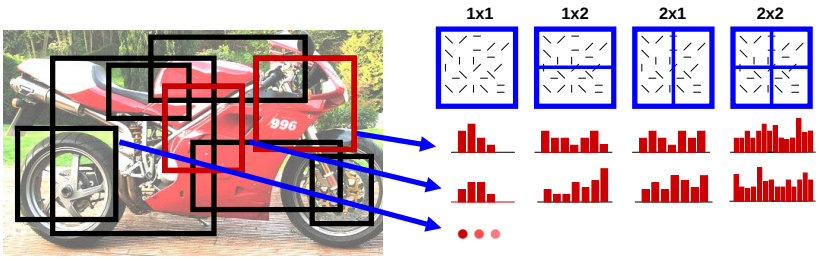

Fig. 3. Histogram features. (Left): Sample regions from an exhaustive set of regions defined by different spatial extents and positions within the object window. (Right): histogram features computed for each region according to four types of spatial grids.

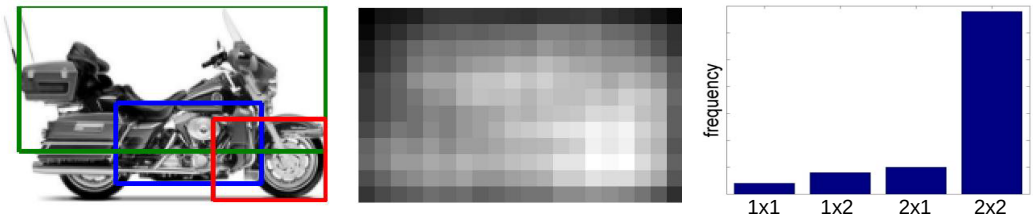

Fig. 4. Selected features for the motorbike class. (Left): Regions of the first three selected and most discriminative features; (Middle): All selected features superimposed using transparent color. Bright areas correspond to the high feature density; (Right): Relative frequency of features with different spatial grids.

3.2 Feature selection

At the training we compute features $f_{k,r}(I)$ for normalised training images and apply AdaBoost to select a set of features $f_{k,r}$ and the corresponding weak classifiers $h(f_{k,r})$ optimizing classification performance. A few features selected for the motorbike class at first rounds of AdaBoost are shown in Figure 4(Left). Superposition of all selected features in Figure 4(Middle) illustrates the emphasis of the final strong classifier on image regions with prominent appearance such as the regions of the front wheel and of the seat. [1] Figure 4(Right) illustrates the high number of selected features with 2x2 spatial grids and indicates the preference of position-dependent histograms for classification.

4 Evaluation

We evaluate the described classifier on the problem of object detection in natural images. To train the classifier for a particular object class, we use positive training set with scale and position-normalised images of objects in similar views. We obtain new negative training samples for each training cascade by collecting false positive detections from training images. For the detection we use the standard window scanning technique and apply the classifier to the large number of image sub-windows with densely sampled positions and sizes. To suppress multiple detections we cluster detected image windows with respect to their positions and sizes in the image and use the size of resulting clusters as a confidence measure of detections.

[1] The asymmetry of selected features in Figure 4(Middle) is explained by the right-alignment of all motorbike image samples used for training.

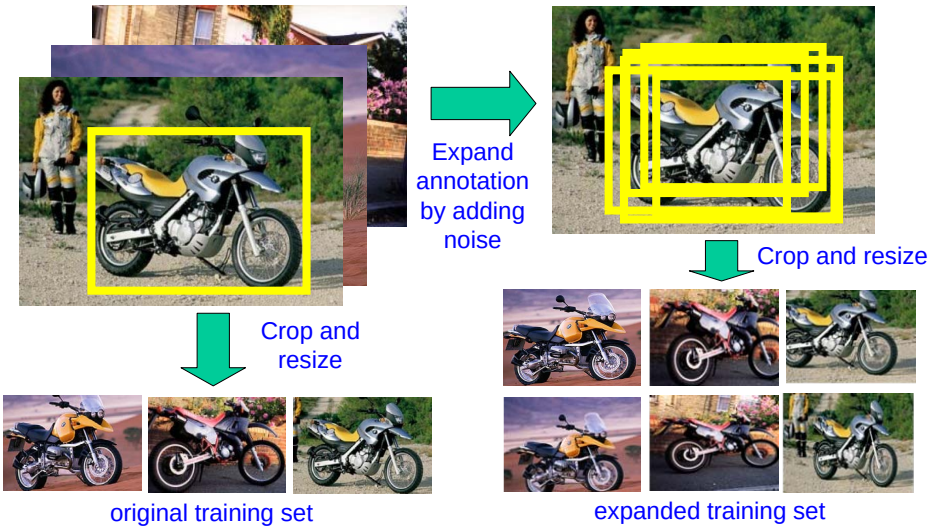

Fig. 5. (Left): Positive training samples are obtained using crop-and-resize procedure applied to training images with rectangular objects annotations. (Right): The same procedure is applied to training images using the large number of automatically generated noisy annotations. Note how annotation noise adds simulated affine deformations to the novel training samples.

To overcome the frequently limited number of positive training samples, we found it particularly useful to artificially enlarge the positive training set as follows. Given annotation rectangles for objects in training images, we generate similar rectangles for each annotation by adding noise to the position and the size of original rectangles. We use noisy annotation to generate new positive image samples and in this way enlarge the positive training set. The procedure is illustrated in Figure 5.

Comparison to Levi and Weiss [15]. Our method differs from the one proposed by Levi and Weiss [15] in three main respects: (i) we introduce WFLD weak learner for vector-valued features, (ii) we use position-dependent histogram features and (iii) we artificially enlarge the positive training set. To evaluate these extensions we compare our method with [15] on the problem of detecting motorbikes in natural images. To train and to test the detectors we use training and validation sets of VOC 2005 challenge and adopt VOC evaluation procedure [7].

Precision-recall evaluation in Figure 6(Left) illustrates gradual improvement of the method in [15] (*1-bin*) with our extensions in terms of *wfld*, position-dependent histograms (*grid*) and the 16 times enlarged training set (*x16*). In addition to the improved performance, *wfld* results in a more efficient classifier with about 25% less features compared to *1-bin*. Surprisingly the largest improvement comes from the increased training set. We further compare the effect of different training set sizes in Figure 6(Right) using *1-bin* classifier.

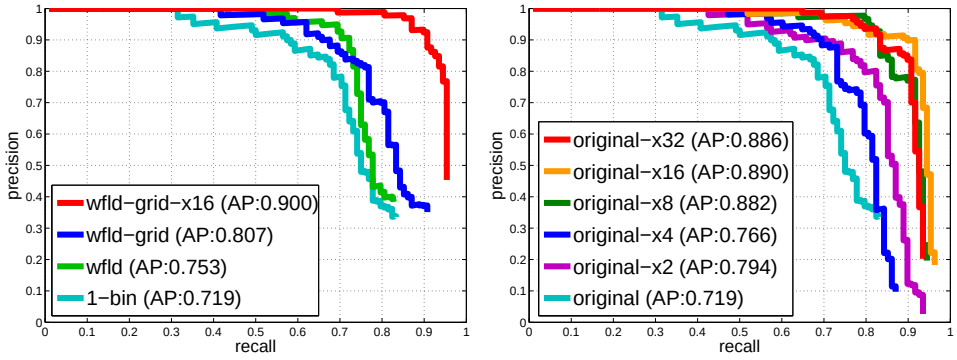

Fig. 6. Detection results for motorbikes on VOC 2005 validation set in terms of Precision-Recall curves and Average Precision (AP) values. (Left): Evaluation of improvements introduced in this paper compared to "1-bin" method [15]; (Right): Evaluation of artificially enlarged training sets.

Comparison on VOC 2005 and VOC 2006 benchmarks. We evaluate the proposed method on PASCAL VOC 2005 and VOC 2006 challenges [7,6] on the task of detecting selected object classes: motorbikes, bicycles, people, cars, horses and cows. The training and the test sets contain substantial variation of objects in terms of scale, pose occlusion and within-class variability. For comparison we select methods with the best detection performance reported in [7,6]. These include (i) *INRIA-Dalal/INRIA_Douze* based on HOG features and linear SVM [2], (ii) *TU-Darmstadt* based on interest points, ISM and SVM [9], (iii) *Edinburgh* using interest points and logistic regression and (iv) *TKK* using image segments and SOM [28].

In Figure 7 and Tables 1,2 our method (*boosted histograms*) demonstrates best results in seven out of eight detection tasks of VOC 2005. The few parameters of our detector (e.g. the number of gradient orientation bins $m = 4$ and the scale of Gaussian derivatives $\sigma = 1$) were optimised on the motorbike validation set and were fixed for the rest of the evaluation. Notably, boosted histograms greatly outperform results reported in [7] for people and bicycles. For motorbikes and cars our method has comparable performance to the best results reported by *INRIA-Dalal* [2] and *TU-Darmstadt* [9]. Note also the difference in relative performance of [2,9] on these two classes and the stable corresponding performance of our method.

Figure 8 shows examples of detection results for motorbikes and people. In Figure 8(Top) the gradual decrease of detection confidence is consistent with the increasing complexity of detected motorbikes. The frequent presence of bicycles within false positives is also intuitive. The detection performance for people is lower in Figure 8(Bottom), however, many high confident false detections (red rectangles) overlap with people in test images. These detections are classified as false positives due to the insufficient overlap with ground truth (green rectangles) or due to the missing annotation.

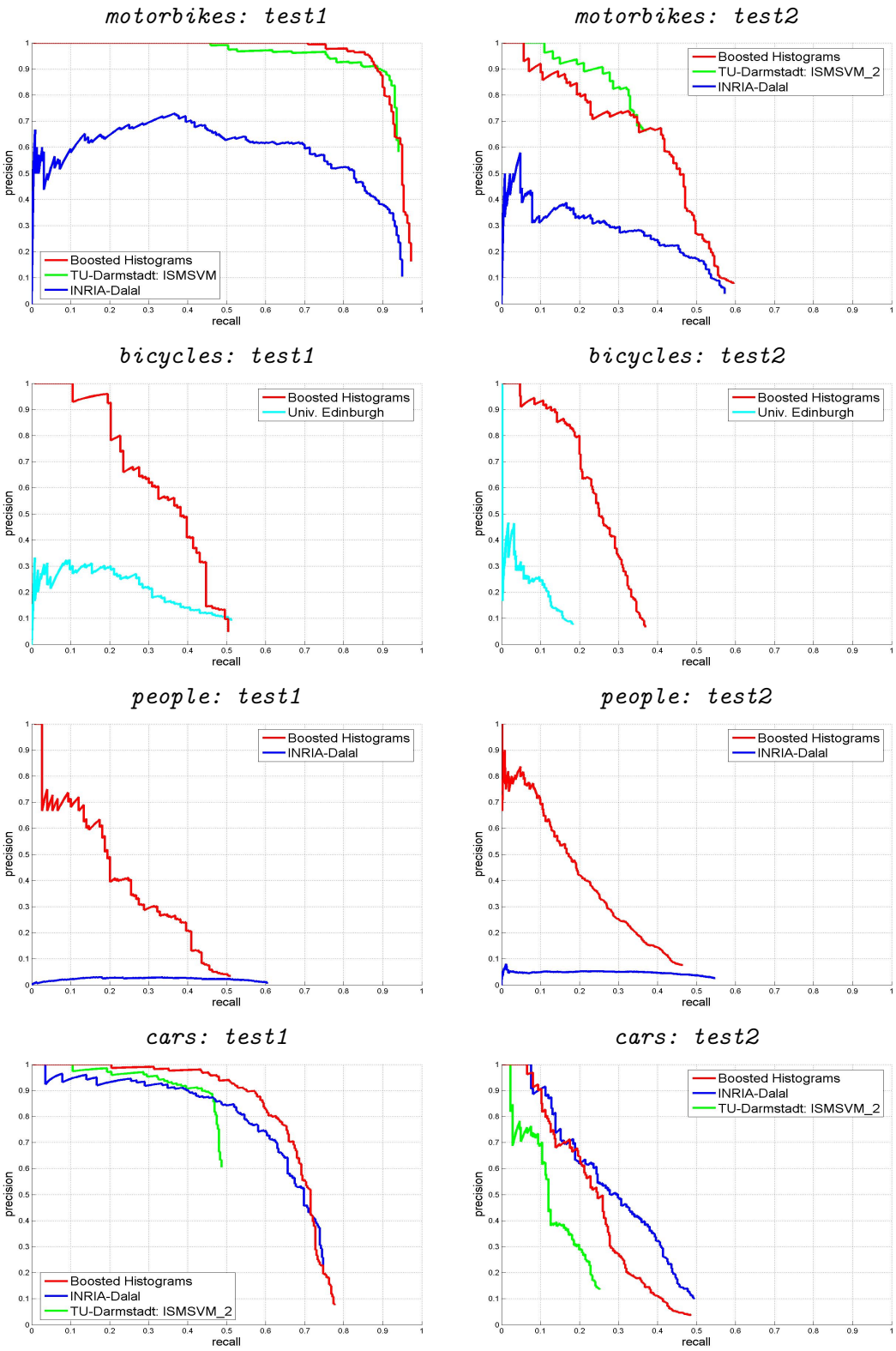

Fig. 7. PR-curves for eight object detection tasks in PASCAL VOC 2005 Challenge. The proposed method (Boosted Histograms) is compared to the best performing methods reported in [7] (better viewed in colour).

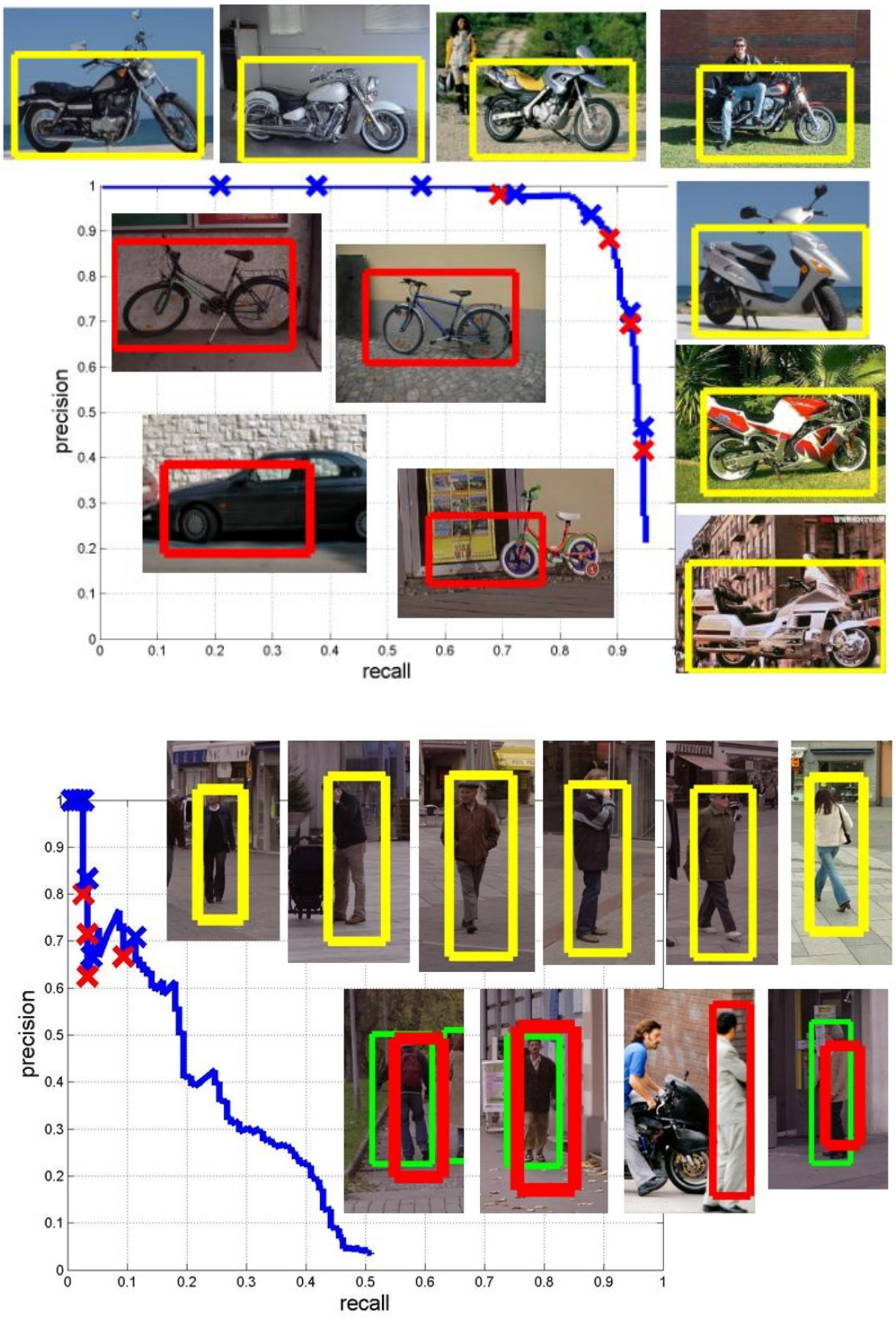

Fig. 8. Examples of true positive (TP) and false positive (FP) detections of motorbikes and people. The location of illustrated detections on PR-curves is marked with crosses. (Top): FP motorbike detections (red) frequently correspond to bicycles. (Bottom): many high confident person detections overlap with people in test images but are frequently classified as FP due to the insufficient overlap with (partly missing) annotation rectangles (green). Better viewed in colour.

Method	Motorbikes	Bicycles	People	Cars
Boosted Histograms	**0.896**	**0.370**	**0.250**	**0.663**
TU-Darmstadt	0.886	–	–	0.489
Edinburgh	0.453	0.119	0.002	0.000
INRIA-Dalal	0.490	–	0.013	0.613

Table 1. Average precision for object detection on `test1` VOC 2005 image set.

Method	Motorbikes	Bicycles	People	Cars
Boosted Histograms	**0.400**	**0.279**	**0.230**	0.267
TU-Darmstadt	0.341	–	–	0.181
Edinburgh	0.116	0.113	0.000	0.028
INRIA-Dalal	0.124	–	0.021	**0.304**

Table 2. Average precision for object detection on `test2` VOC 2005 image set.

Method	bicycle	cow	horse	motorbike	person
INRIA_Douze	0.414	0.212	–	**0.390**	**0.164**
INRIA_Laptev	**0.440**	0.224	**0.140**	0.318	0.114
TKK	0.303	**0.252**	0.137	0.265	0.039

Table 3. Average precision for object detection in `task 3` of VOC 2006.

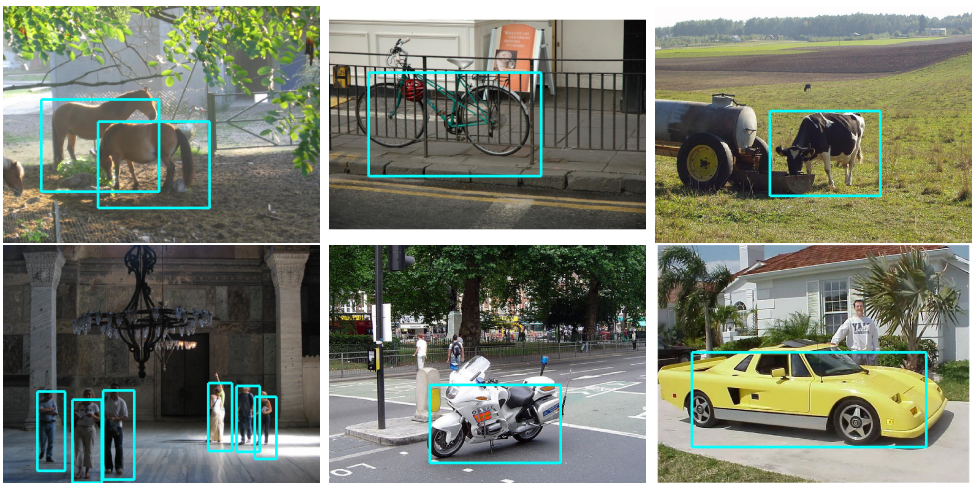

Fig. 9. Detection results for horses, bicycles, cows, people, motorbikes and cars.

The PASCAL VOC Challenge 2006 [6] contains ten object classes. We actively participated in the challenge and submitted results for five object classes: *bicycle*, *cow*, *horse*, *motorbike* and *person*. Among these five classes the boosted histogram method obtained best results for classes *bicycle* and *horse* while second-best results were obtained for three other object classes as summarised in Table 3. Example detections for a few test images are illustrated in Figure 9.

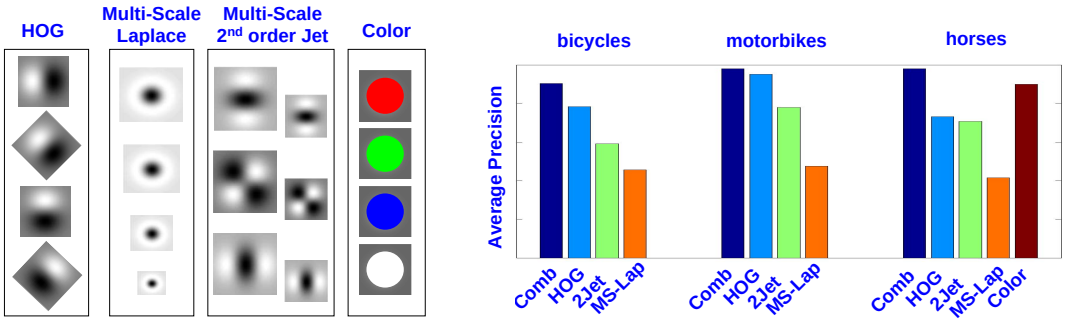

Fig. 10. Alternative histogram features. (Left): Filter banks used to construct four types of histogram features in this paper. Each filter corresponds to one bin of the histogram. (Right): Relative performance of different histogram features and their combinations applied to the detection of three object classes.

5 Alternative image features

In this section we consider alternative histogram features and feature selection mechanisms and evaluate our method augment with such extensions on object detection tasks.

5.1 Histograms of color and second order image derivatives

We investigate whether the histograms of alternative image properties can provide better or complementary performance with respect to HOG features. For this purpose in addition to HOGs we introduce three histogram descriptors defined by local image measurements in terms of (i) multi-scale Laplacian responses, (ii) second order jet responses and (iii) color as illustrated in Figure 10(Left). The choice of Laplacian features is motivated by their rotation invariance and scale selection property [17,18]. Second order jets [11] capture local second order differential image structure while color is discriminative e.g. for certain animal classes. To construct histograms, we maximize responses over associated filters at every image point and increment corresponding histogram bins. The training and the detection then follows the same procedure as described for HOG features in previous sections.

Relative performance of different histogram features in Figure 10(Right) illustrates superior performance of HOG compared to other gray-scale descriptors. Color histograms outperform HOG for horses but result in poor training convergence for other two object classes. The best performance for all tested classes is achieved by the combination of all features. The combination was achieved by clustering multiple responses of alternative detectors trained separately for each type of features.

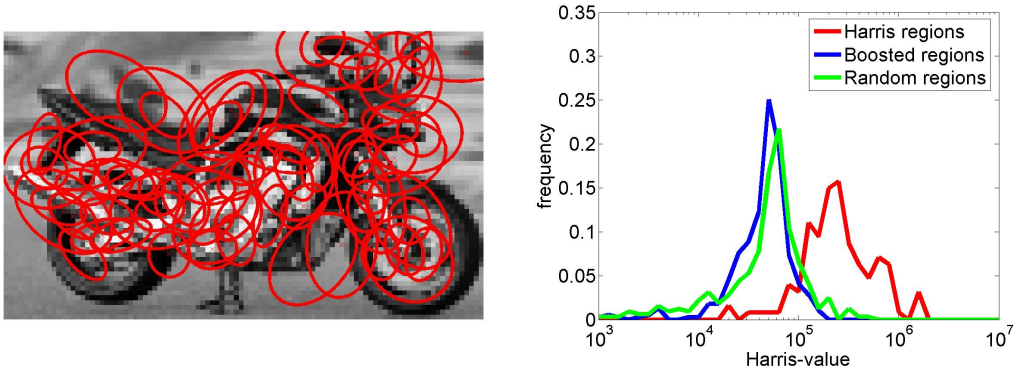

Fig. 11. Comparison of boosted histogram regions with Harris-Affine features [20]. (Left): Harris-Affine regions detected for a motorbike image. (Right): Distribution of Harris values for different types of features.

5.2 Alternative feature selection

Interest point features have been a popular choice of local image descriptors in many recognition methods [9,19,25,31]. We investigate if these descriptors bear similarity with histogram features selected by our method. For this purpose we choose Harris-Affine features [20] as an example of a popular region detector illustrated in Figure 11(Left). We then compare the values of Harris function [10] computed for Harris-Affine features, boosted regions and random regions on motorbike images. Distributions of Harris values for these three types of regions are illustrated in Figure 11(Right). As expected, the responses of Harris function are higher for Harris-Affine features compared to random regions. Notably, boosted regions show low responses for Harris function and, hence, bear low similarity to Harris-Affine features.

We next investigate whether the boosted histogram detector can be improved by using Harris interest regions for training. For this purpose we pre-select fractions of features using (a) random selection of regions and (b) selection of interest regions maximising the Harris function. We train classifiers for three VOC 2005 object classes using different fractions of pre-selected features and different selection methods and evaluate the detection performance in Figure 12(Top). Notably, the performance of random features is similar or better compared to Harris features. At the same time the complexity of classifiers trained on Harris regions is higher compared to random regions according to Figure 12(Bottom). This indicates that image features selected by the popular Harris function may not always be the best choice in a recognition system.

In Figure 12 we also observe the very stable performance and complexity of detectors trained on 10% randomly selected regions only. This implies the opportunity to speed up the training procedure without penalizing the performance and the complexity of the detection.

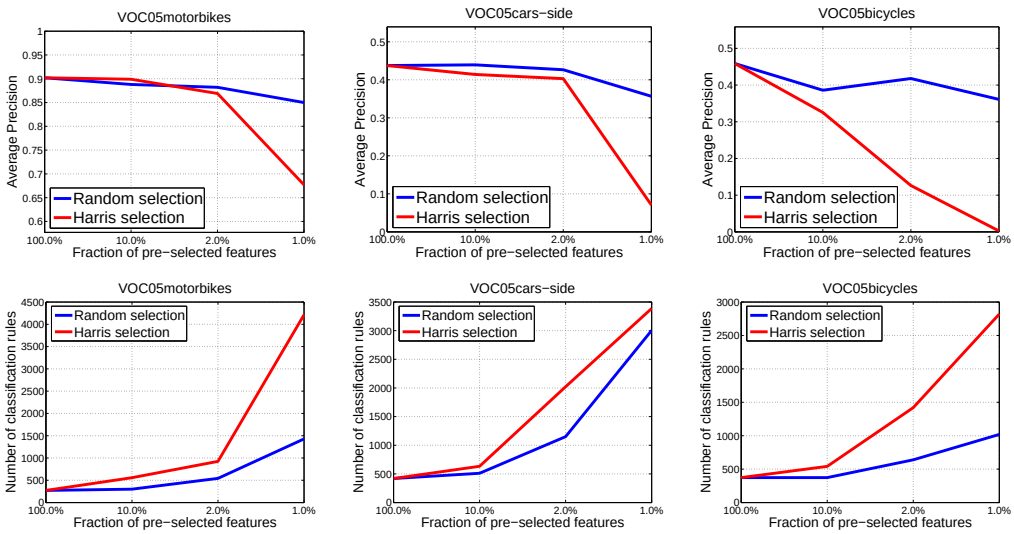

Fig. 12. (Top): Detection performance of boosted histogram detectors trained on different fractions of regions selected either randomly or by maximizing responses of the Harris function. (Bottom): Complexity of corresponding detectors measured by the number of selected features. (Better viewed in colour.)

6 Computational aspects

In their face detection method Viola and Jones [29] introduced integral images for the fast computation of rectangular grey-level features. This idea was further developed to integral histograms [15,22] to enable fast computation of histograms in rectangular image regions of arbitrary positions and sizes. A major difference to the original approach in [29] arises, however, when computing histograms of filter responses for multi-scale tasks such as for object detection at multiple image resolutions.

Filter responses such as the responses of Gaussian derivatives are known to change over image scales [16]. Hence, to enable unbiased computation of histograms at different scales, either the size of filter kernels or the image resolution has to be adapted to the scale parameter. This, however, implies additional computational cost due to a separate filtering step and the recomputation of integral histograms at each scale level.

Given the high correlation of filter responses at adjacent image scales, computation of integral histograms for a limited set of sparse scale levels is likely to imply a speed up at the cost of a limited decrease of performance. To investigate this issue in the context of our detection algorithm we introduce the following parameters. We denote the number of scale levels in octave by α implying a scale factor of $2^{1/\alpha}$ between adjacent scale levels. We recompute integral histograms at each β^{th} scale level $c = n\beta, n \in \mathbb{Z}$ only. At other scale levels c_i we accommodate for scale changes by resizing rectangular features of

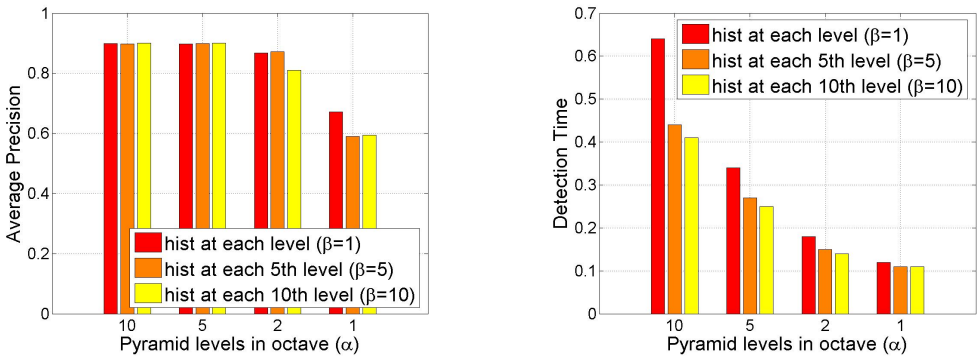

Fig. 13. Precision-speed tradeoff. Average precision values (Left) and the detection speed (Right) are illustrated for different densities of scale sampling (α) and different scale steps of integral histograms (β). Higher speed of detection can be achieved by means of a coarse scale sampling without compromising precision.

the object classifier similar to [29] while deriving histogram features from an integral histogram at scale level $c = \beta\lfloor c_i/\beta \rfloor$.

To study the tradeoff between the speed and the accuracy of our detection method we perform a set of experiments using different values of parameters α and β while measuring average precision of detection on the VOC05 motorbike validation dataset. As illustrated in Figure 13(left) the precision of detection remains stable for $\alpha = 5, 10$ and $\beta = 1, 5, 10$ while the detection speed increases more than twice (see Figure 13,right). Choosing two scale levels in octave ($\alpha = 2$) while recomputing integral histograms at each 5^{th} scale level only ($\beta = 5$) seems to give a near optimal precision-speed tradeoff on this dataset. We have observed similar behaviour for detectors trained on other object classes. Our current implementation of object detection runs at about 10fps frame rate on 320×240 images on a modest PC. The implementation source code is available for download [2] .

7 Conclusion

We presented a method for object detection that combines AdaBoost learning with local histogram features. While being conceptually similar to [15] our method provides a number of extensions that significantly improve the results of object detection. We evaluated the method on recent benchmarks for object recognition [7,6] and demonstrated its competitive performance compared to the state-of-the-art. We also addressed computational aspects of the method by analysing precision-seed tradeoff of detection.

[2] `http://www.irisa.fr/vista/Equipe/People/Laptev/objectdetection.html`

Acknowledgements

The author would like to thank Patrick Pérez and Patrick Bouthemy for their helpful comments. Mark Everingham, Mario Fritz and Navneet Dalal were extremely helpful providing details and results of the VOC 2005 Challenge.

References

[1] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (4) (2002) 509–522.

[2] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proc. Computer Vision and Pattern Recognition, 2005.

[3] G. Dorkó, C. Schmid, Selection of scale-invariant parts for object class recognition, in: Proc. Ninth International Conference on Computer Vision, Nice, France, 2003.

[4] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, 2001.

[5] B. Epshtein, S. Ullman, Semantic hierarchies for recognizing objects and parts, in: Proc. Computer Vision and Pattern Recognition, 2007.

[6] M. Everingham, L. Van Gool, C. Williams, A. Zisserman, The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results, http://www.pascal-network.org/challenges/VOC/voc2006.

[7] M. Everingham, A. Zisserman, C. Williams, L. Van Gool, M. Allan, C. Bishop, O. Chapelle, N. Dalal, T. Deselaers, G. Dorko, S. Duffner, J. Eichhorn, J. Farquhar, M. Fritz, C. Garcia, T. Griffiths, F. Jurie, D. Keysers, M. Koskela, J. Laaksonen, D. Larlus, B. Leibe, H. Meng, H. Ney, B. Schiele, C. Schmid, E. Seemann, J. Shawe-Taylor, A. Storkey, S. Szedmak, B. Triggs, I. Ulusoy, V. Viitaniemi, Z. J., The 2005 pascal visual object classes challenge, in: Selected Proceedings of the First PASCAL Challenges Workshop, 2005.

[8] Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences 55 (1) (1997) 119–139.

[9] M. Fritz, B. Leibe, B. Caputo, B. Schiele, Integrating representative and discriminative models for object category detection, in: Proc. 10th International Conference on Computer Vision, Beijing, China, 2005.

[10] C. Harris, M. Stephens, A combined corner and edge detector, in: Alvey Vision Conference, 1988.

[11] J. Koenderink, A. van Doorn, Representation of local geometry in the visual system, Biological Cybernetics 55 (1987) 367–375.

[12] I. Laptev, Improvements of object detection using boosted histograms, in: Proc. British Machine Vision Conference, Edinburgh, UK, 2006.

[13] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proc. Computer Vision and Pattern Recognition, 2006.

[14] T. Leung, J. Malik, Representing and recognizing the visual appearance of materials using three-dimensional textons, International Journal of Computer Vision 43 (1) (2001) 29–44.

[15] K. Levi, Y. Weiss, Learning object detection from a small number of examples: The importance of good features, in: Proc. Computer Vision and Pattern Recognition, 2004.

[16] T. Lindeberg, Scale-Space Theory in Computer Vision, Kluwer Academic Publishers, Boston, 1994.

[17] T. Lindeberg, Feature detection with automatic scale selection, International Journal of Computer Vision 30 (2) (1998) 77–116.

[18] D. Lowe, Object recognition from local scale-invariant features, in: Proc. Seventh International Conference on Computer Vision, Corfu, Greece, 1999.

[19] K. Mikolajczyk, B. Leibe, B. Schiele, Local features for object class recognition, in: Proc. 10th International Conference on Computer Vision, Beijing, China, 2005.

[20] K. Mikolajczyk, C. Schmid, An affine invariant interest point detector, in: Proc. Seventh European Conference on Computer Vision, vol. 2350 of Lecture Notes in Computer Science, Springer Verlag, Berlin, Copenhagen, Denmark, 2002.

[21] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, in: Proc. Computer Vision and Pattern Recognition, 2003.

[22] F. Porikli, Integral histogram: A fast way to extract histograms in cartesian spaces, in: Proc. Computer Vision and Pattern Recognition, 2005.

[23] B. Schiele, J. Crowley, Recognition without correspondence using multidimensional receptive field histograms, International Journal of Computer Vision 36 (1) (2000) 31–50.

[24] H. Schneiderman, T. Kanade, A statistical method for 3D object detection applied to faces and cars, in: Proc. Computer Vision and Pattern Recognition, vol. I, Hilton Head, SC, 2000.

[25] J. Sivic, B. Russell, A. Efros, A. Zisserman, W. Freeman, Discovering objects and their localization in images, in: Proc. 10th International Conference on Computer Vision, Beijing, China, 2005.

[26] M. Swain, D. Ballard, Color indexing, International Journal of Computer Vision 7 (1) (1991) 11–32.

[27] M. Varma, A. Zisserman, Classifying images of materials: Achieving viewpoint and illumination independence, in: Proc. Seventh European Conference on Computer Vision, Lecture Notes in Computer Science, Springer Verlag, Berlin, Copenhagen, Denmark, 2002.

[28] V. Viitaniemi, J. Laaksonen, Techniques for still image scene classification and object detection, in: 16th Int. Conf. on Artificial Neural Networks (ICANN), 2006.

[29] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proc. Computer Vision and Pattern Recognition, Kauai Marriott, Hawaii, 2001.

[30] H. Wang, P. Li, T. Zhang, Histogram features-based fisher linear discriminant for face detection, in: Proc. Asian Conference on Computer Vision, 2006.

[31] J. Zhang, M. Marszalek, S. Lazebnik, C. Schmid, Local features and kernels for classification of texture and object categories: A comprehensive study, International Journal of Computer Vision 73 (2) (2007) 213–238.

[32] Q. Zhu, M. Yeh, K. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in: Proc. Computer Vision and Pattern Recognition, 2006.