Improvements of Object Detection Using Boosted Histograms

Ivan Laptev
IRISA / INRIA Rennes
35042 Rennes Cedex France
ivan.laptev@inria.fr

Abstract

We present a method for object detection that combines AdaBoost learning with local histogram features. On the side of learning we improve the performance by designing a weak learner for multi-valued features based on Weighted Fisher Linear Discriminant. Evaluation on the recent benchmark for object detection confirms the superior performance of our method compared to the state-of-the-art. In particular, using a *single* set of parameters our approach outperforms *all* methods reported in [5] for 7 out of 8 detection tasks and four object classes.

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 [16] proposed color histograms as an early view-based method for object recognition. The idea was further developed by Schiele and Crowley [14] who recognized objects using histograms of local filter responses. Histograms of Textons were proposed by Leung and Malik [8] for texture recognition. Schneiderman and Kanade [15] computed histograms of wavelet coefficients over localized object parts and were among the first to address object categorization in natural scenes. In a similar spirit the well-known SIFT descriptor [10] and Shape Context [1] use position-dependent histograms computed in the neighbourhood of selected image points.

Histograms represent distributions of spatially unordered image measurements in a region and provide relative invariance to several object transformations in the image. This property partly explains the success of histogram-based methods. The invariance and the descriptive power of histograms, however, crucially depend on (a) the type of local image measurements and (b) the image region used to accumulate histograms. Regarding the type of measurements, different alternatives have been proposed that may have better performance depending on the recognition task [16, 14]. As a general purpose shape descriptor, the choice of histograms of gradient orientations is well supported by many applications of SIFT descriptor [10, 12] and other related methods [2].

Besides the question *what* to measure, the question *where* to measure obviously has a large impact on recognition. While global histograms [16, 14] do not suite well for complex scenes, a better approach supported in [15, 10, 2] consists of computing histograms over local image regions. As illustrated in Figure 1, different regions of an object may

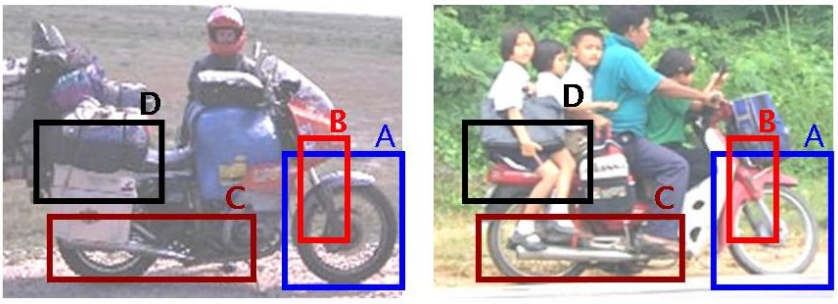

Figure 1: Rectangles on the left and right image are examples of possible regions for histogram features. Stable appearance in A,B and C on both images makes corresponding features to be good candidates for a motorbike classifier. On the contrary, regions D are unlikely to contribute for the classification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

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 by the tessellation [15, 2] or by applying region detectors of different kinds [10, 3, 11]. While many region detectors were designed to achieve invariance to local geometric transformations, it should be stressed that the procedures used to detect such regions are based on heuristic functions[1] and cannot guarantee optimal recognition. An arguably more attractive alternative proposed by Levi and Weiss [9] consists of learning class-specific histogram regions from the training data.

In this work similar to [9] we choose the position and the shape of histogram features to minimize the training error for a given recognition task. We consider a complete set of rectangular regions in the normalized object window and compute histograms of gradient orientation for several parts of such regions. We then apply AdaBoost procedure [6, 18] to select histogram features (Boosted Histograms) 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). This together with other improvements is shown to substantially improve the performance of the method in [9].

As our second contribution, we apply the developed method to the problem of object detection in cluttered scenes and evaluate the performance on the benchmark of PASCAL Visual Object Category (VOC) Challenge 2005 [5]. Using a *single* set of parameters our approach outperforms *all* methods reported in [5] for 7 out of 8 detection tasks and four object classes. Among the advantages of the method we reinforce and emphasize (a) its ability to learn from a small number of samples, (b) stable performance for different object classes, (c) conceptual simplicity and (d) potentially real-time implementation.

The rest of the paper is organized 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 apply the method to object detection and evaluate its performance. Section 5 concludes the paper.

[1] For example Harris function for position estimation and the normalized Laplacian for scale selection.

2 AdaBoost learning

AdaBoost [6] 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

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

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 maximizing the classification performance. This idea was first explored by Viola and Jones [18] who used AdaBoost to train an efficient face detector by selecting a discriminative set of local Haar features. Here similar to [9], 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 [18], an efficient classifier for n training samples can be found by selecting an optimal decision threshold θ 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 [9] 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 $g_2(f) = (0\ 1)f$, the resulting AdaBoost classifier has poor generalization and requires $T \approx 50$ weak hypotheses for separating $n = 200$ training samples.

An alternative and still efficient choice for a multi-dimensional classifier is Fisher Linear Discriminant (FLD) [4]. FLD guarantees optimal classification of normally distributed samples of two classes using a linear projection function

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

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

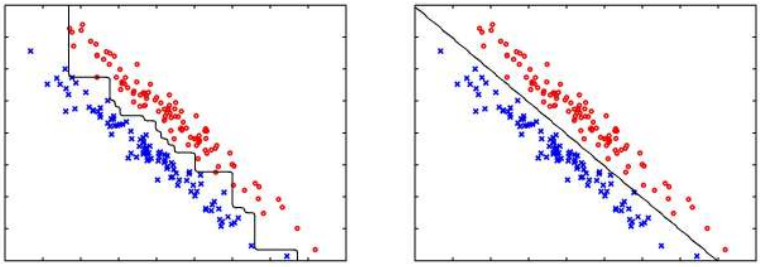

Figure 2: Classification of two diagonal distributions using (left): AdaBoost with weak learners in terms of axis-parallel linear classifiers; (right): Fisher linear discriminant.

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 minimize 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 the training data required for other classifiers that do not accept weighted samples. This in turn leads to a more efficient use of the training data which is frequently limited in vision applications.

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 resulting classifier. Experimental validation of this assumption and the advantage of WFLD will be demonstrated in Section 4 on real classification problems.

3 Histogram features

As motivated in the introduction, local histograms provide effective means to represent visual information for recognition. To avoid a-priori selection of histogram regions, we consider all rectangular sub-windows r of the object. For image regions r we compute weighted histograms of gradient orientations

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

using Gaussian derivatives L_x, L_y defined on the image I. We discretize γ into $m = 4$ orientation bins and increment histograms by the values of the gradient magnitude $||(L_x^2, L_y^2)||_2$. The histograms are normalized to the sum value 1 to reduce the influence of illumination.

To preserve some positional information of measurements within the region, we subdivide regions into parts as illustrated in Figure 3(upper,left) and compute histograms

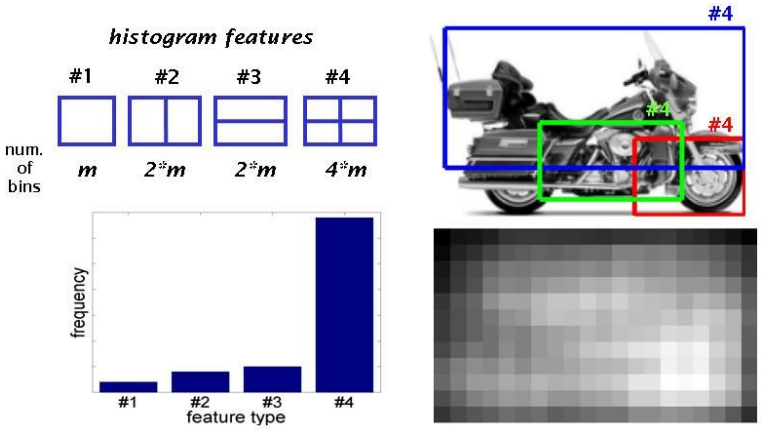

Figure 3: (top,left): Four types of compound histogram features; (bottom,left): Frequency of the types of compound features in the AdaBoost motorbike classifier; (top,right): Features chosen in the three first rounds $t = 1,2,3$ of AdaBoost learning; (bottom,right): Superposition of all rectangular features selected for a motorbike classifier. The value at each pixel corresponds to the number of selected regions that overlap with the pixel.

separately for each part. Four *types* of image features $f_{k,r}(z),\ k = 1,...,4$ are then defined for each region r by concatenating part-histograms into feature vectors of dimensions $m, 2m, 2m$ and $4m$ respectively. All histogram features are computed efficiently using integral histograms [9, 13] which enables real-time implementation of the detection method.

During training we compute features $f_{k,r}(z)$ for the normalized training images and apply AdaBoost to select a set of features $f_{k,r}$ and hypotheses $h(f_{k,r})$ for optimal performance of classification. A few features selected for motorbikes in the first rounds of AdaBoost are shown in Figure 3(upper,right). By superimposing the regions of all selected features illustrated in Figure 3(lower,right) we can observe the relative importance of different parts of the motorbike for the classification. The frequency of selected feature types is illustrated in Figure 3(lower,left) and indicates the preference of compound features for the classification.

4 Evaluation

We evaluate the designed classifier on the problem of object detection in natural scenes. For the training we assume a set of scale and position normalized object images with similar views. We use a cascade AdaBoost classifier [18] and collect negative examples for training by detecting false positives in random images. For the detection we use the standard window scanning technique and apply the classifier to the densely sampled subwindows of the image. To suppress multiple detections we cluster positively classified sub-windows in the position-scale space and use the size of resulting clusters as a confidence measure for the detection.

To improve the performance of object detection, we found it particularly useful to populate the training set of positive samples as follows. Given annotation rectangles for objects in training images, we generate similar rectangles for each annotation by perturbing the position and the size of original rectangles. We treat the generated rectangles as new annotations and populate the training set of positive samples by the factor of 10.

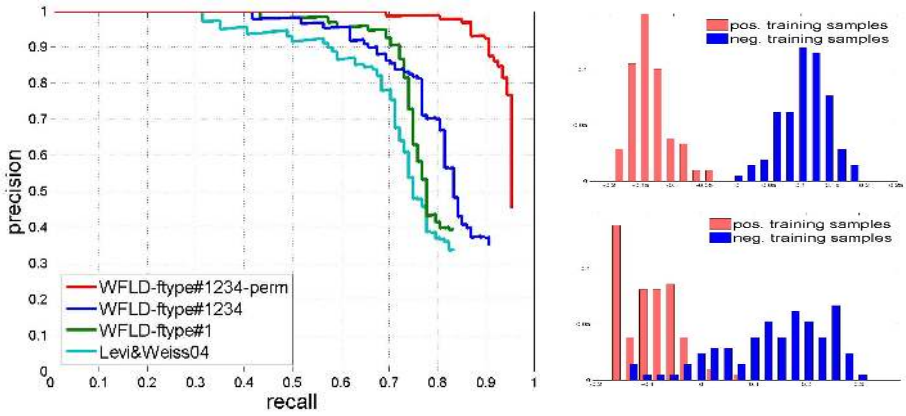

Figure 4: (Left): Comparison of detection methods using Precision-Recall curves. (Right): Distributions of training samples projected onto examples of basis functions selected by different weak learners (top): WFLD; (bottom): Levi&Weiss04.

Comparison to Levi and Weiss [9]. Our method differs from the one proposed by Levi and Weiss [9] in three main respects: (i) we introduce WFLD weak learner for vector-valued features, (ii) we use compound histogram features described in Section 3 and (iii) we use a populated set of training samples as described above. To evaluate these extensions we compare our method with [9] on the problem of detecting motorbikes in natural scenes. To train and to test the detectors we used the training and the validation datasets of VOC 2005 [5]. Evaluation in terms of Precision-Recall (PR) curves illustrated in Figure 4(left) shows how the performance of the method in [9] is gradually improved by our extensions. In particular we noticed that WFLD usually gave a better separation of training samples as illustrated in Figure 4(right) as well as resulted in a simpler classifier with about 25% less weak classifiers than required by our implementation of [9]. Surprisingly we observed that most of improvement was given by the population of the training set.

Comparison to VOC 2005. One of our main contributions is the evaluation of the presented method on the VOC 2005 dataset [5]. In [5] several state-of-the-art methods for object detection were evaluated on the problem of detecting four object classes: motorbikes, bicycles, people and cars. The training and the two test sets contained substantial variation of objects in terms of scale, pose occlusion and within-class variability. The evaluation was done in terms of PR curves and the Average Precision (AP) values approximating the area under the PR-curves (see [5] for details).

As follows from Figure 5 and Tables 1,2 our method denoted as *Boosted Histograms* outperforms the best results in [5] in seven out of eight test problems. To generate the results we *did not* optimize our method for each object class. The (few) parameters of our detector such as the number of histogram bins $m = 4$ and the scale of Gaussian derivatives $\sigma = 1$ in (4) were optimized on the validation set of the motorbike class and were fixed for the rest of object classes. Notably, the performance of Boosted Histograms (BH) greatly outperforms results in [5] for people and bicycles. For motorbikes and cars we note that BH performs better or similar to competitor methods [2, 7] while the relative performance of [2] and [7] is rather different for these two object classes.

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.

Figure 6 shows examples of detection results for motorbikes and people. In Figure 6(top) gradual decrease of the detection confidence is consistent with the increased complexity of detected motorbikes. The frequent presence of bicycles within false positives can also be explained intuitively. Moreover, exclusive fusion of detection results for motorbikes and bicycles is expected to increase the detection results for both classes even further. In Figure 6(bottom) we observe that acceptable detections of people (red rectangles) are frequently labelled as "misclassified" during the evaluation due to the misalignment with the annotation (green rectangles) or due to the missing annotation.

5 Conclusion

We presented a method for object detection that combines AdaBoost learning with local histogram features. While being conceptually similar to [9] our method provides a number of extensions that significantly improve the results of object detection. We evaluated the method on the recent benchmark for object detection [5] and demonstrated its superior performance compared to the state-of-the-art methods reported in [5]. Based on the observations in Section 4 we conclude that the current method has a stable performance for different objects classes.

Among the possible extensions, the current method can be easily re-formulated to capture histograms of other image measurements such as textons. This might further improve the performance by adapting the method to particular object classes. On the side of learning more work towards efficient weak learners might be fruitful. Re-formulating the current method for multi-class problems using multi-class version of AdaBoost [17] is another potentially interesting extension.

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.

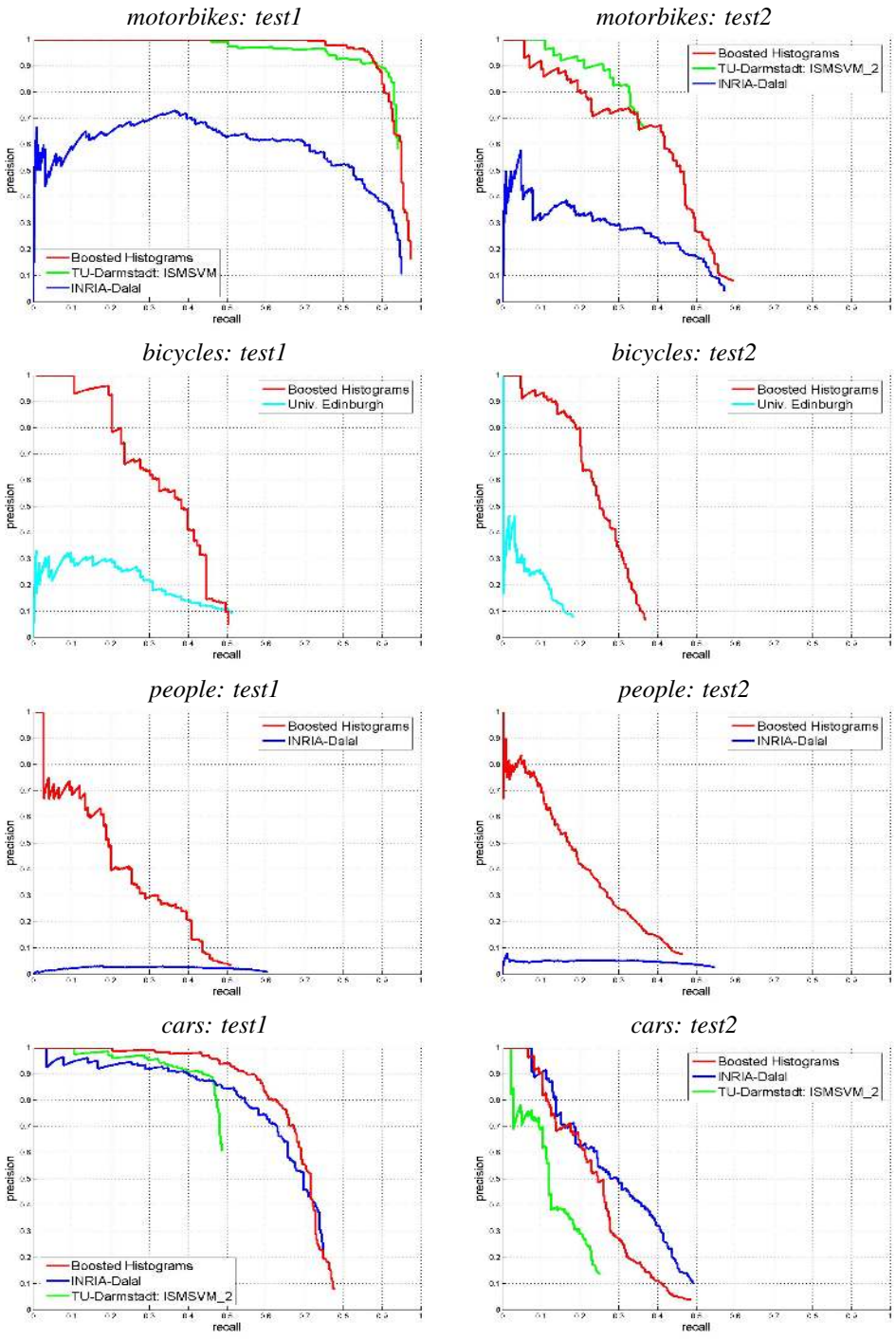

Figure 5: PR-curves for eight object detection tasks of PASCAL VOC 2005 Challenge. The proposed method (Boosted Histograms) is compared to the best detection methods reported for each task in [5]. (This Figure is better viewed in colour.)

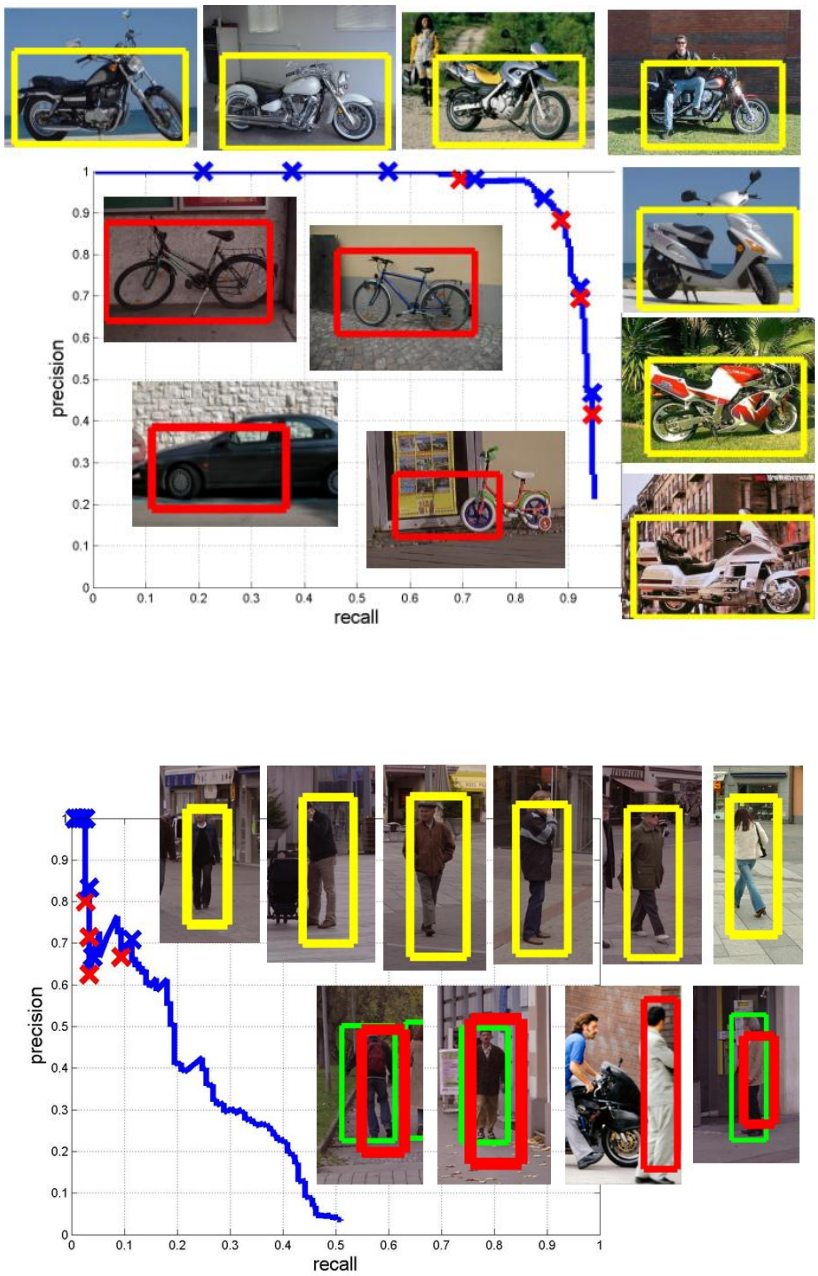

Figure 6: Examples of correct and false detections of motorbikes and people. The positions of illustrated detections on the PR-curves are marked with the crosses. (top): False detections of motorbikes (red rectangles) frequently correspond to bicycles. (bottom): Acceptable detections of people (red rectangles) are frequently labelled as "misclassified" in the evaluation due to the misalignment with the annotation (green rectangles).

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE-PAMI*, 24(4):509–522, April 2002.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, pages I:886–893, 2005.

[3] Gyuri Dorkó and Cordelia Schmid. Selection of scale-invariant parts for object class recognition. In *Proc. ICCV*, pages I:634–640, 2003.

[4] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley, 2001.

[5] 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, and Zhang J. The 2005 pascal visual object classes challenge. In *Selected Proceedings of the First PASCAL Challenges Workshop*, 2005.

[6] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *J. of Comp. and Sys. Sc.*, 55(1):119–139, 1997.

[7] M. Fritz, B. Leibe, B. Caputo, and B. Schiele. Integrating representative and discriminative models for object category detection. In *Proc. ICCV*, pages II:1363–1370, 2005.

[8] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44, June 2001.

[9] K. Levi and Y. Weiss. Learning object detection from a small number of examples: The importance of good features. In *Proc. CVPR*, pages II:53–60, 2004.

[10] D.G. Lowe. Object recognition from local scale-invariant features. In *Proc. ICCV*, pages 1150–1157, 1999.

[11] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Proc. ICCV*, pages II:1792–1799, 2005.

[12] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proc. CVPR*, pages II: 257–263, 2003.

[13] F.M. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *Proc. CVPR*, pages I:829–836, 2005.

[14] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *IJCV*, 36(1):31–50, January 2000.

[15] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. CVPR*, volume I, pages 746–751, 2000.

[16] M.J. Swain and D.H. Ballard. Color indexing. *IJCV*, 7(1):11–32, November 1991.

[17] A. Torralba, K.P. Murphy, and W.T. Freeman. Sharing features: Efficient boosting procedures for multiclass object detection. In *Proc. CVPR*, pages II:762–769, 2004.

[18] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. CVPR*, pages I:511–518, 2001.