Exploiting Object Characteristics Using Custom Features for Boosting-Based Classification

Arne Ehlers, Florian Baumann, and Bodo Rosenhahn

Institut für Informationsverarbeitung (TNT)
Leibniz Universität Hannover, Germany
lastname@tnt.uni-hannover.de

Abstract. Typical feature pools used to train boosted object detectors contain various redundant and unspecific information which often yield less discriminative detectors. In this paper we introduce a feature mining algorithm taking domain specific knowledge into account. Our proposed feature pool contains rectangular shaped features generated from an image clustering algorithm applied on the mean image of the object training set. A combination of two such spatially separated rectangular regions yields a set of features which have a similar evaluation time like classical Haar-like features, but are much smarter (automatically) selected and more discriminative since image correlations can be more consequently exploited. Overall, training is faster and results in more selective detectors showing improved precision. Several experiments demonstrate the gain when using our proposed feature set in contrast to standard features.

1 Introduction

An object detection framework that utilizes machine learning in a training phase commonly requires a positive and negative object set as well as a set of features to create the classifier. The features are required to correlate with the object class such that the machine learning algorithm can make use of the feature response to distinguish between the object classes. Various object detection frameworks inspired by Viola and Jones [1] are based on simple Haar-like features that describe the difference in intensity between rectangular image regions.

Starting from a small set of basis features, a complete feature pool is usually created by translation and scaling of these basis features. This procedure makes the number of possible features highly dependent on the size of the images in the training set. The derived feature pool commonly contains a very large number of irrelevant and redundant features which are often less discriminative. Furthermore, these approaches to construct a feature pool completely neglect the available domain knowledge provided by the positive training set.

1.1 Related Work

The task of extracting knowledge from such a data set is addressed by data mining. In terms of the field of data mining, a distinction can be made between feature selection, i.e. create a subset by choosing reasonable features from a large feature pool and feature

J.-K. Kämäräinen and M. Koskela (Eds.): SCIA 2013, LNCS 7944, pp. 420–431, 2013.

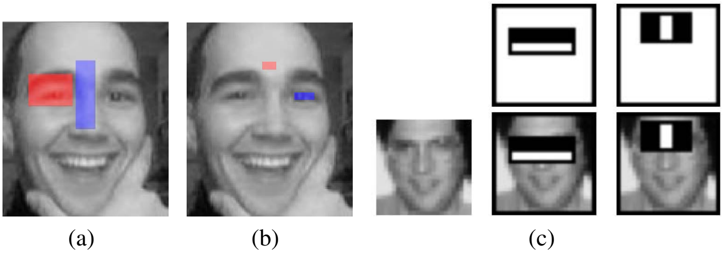


Fig. 1. Demonstration of chosen uncontiguous 2Rec-Features and Viola and Jones first and second chosen Haar-like feature. [1].

mining, i.e. construct an own complete feature pool [2]. Blum and Langley [3] categorize feature selection methods in three parts: (1) embedded, i.e. the well-known feature selection mechanism by Viola and Jones [1]. (2) filter, performing a separate process such as filter out irrelevant information and (3) wrapper, reapplying the trained classifier to the subset with additional computational cost. Koller and Sahami [4] proposed a method for selecting a subset of features by eliminating features providing little or no additional information. But extracting a feature pool directly from the data set yields usually to a better performance and a more discriminative power [5].

1.2 Contribution

Inspired by this field of research we propose a new feature type along with its feature mining scheme, in the following referred to as 2Rec-Feature, which can describe spatially separated regions (see Figure 1(a) and 1(b)) and overcomes the adjacency constraint of conventional Haar-like features (see Figure 1(c)). An automatic feature-creation process exploits the characteristics of the object class and enables a much richer representation and variability in the feature set. Instead of deriving a feature pool from general basis features, all 2Rec-Features are directly determined in position and scale with respect to intrinsic properties of the object class. Without this derivation from feature mining, the added variability of spatially separated regions would lead to a substantially larger feature set. Additional hard constraints on position and size of the features would have to be introduced to render the feature set manageable in the training phase as Zhao et al. implemented in [6]. In contrast, we eliminate the dependency of the feature pool size on the training image size and directly construct a small but representative feature set to enable a faster classifier training.

We evaluate our new feature class on the well-established MIT+CMU upright face test-set [7], see Figure 2(a), on the Face Detection Data Set and Benchmark for unconstrained face detection (FDDB) [8], see Figure 2(b), and on UIUC lateral car-detection [9], see Figure 2(c).

The paper is structured as follows. Section 2 presents our proposed feature class. In Section 3 the applied machine learning framework is briefly described. Experimental results comparing 2Rec-Features to conventional Haar-like features and state-of-the-art methods are presented in Section 4. Section 5 summarizes and concludes the paper.

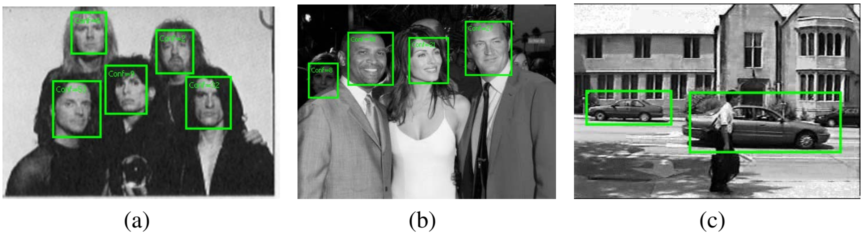


Fig. 2. Example detections of our method applied on the MIT+CMU [7], FDDB [8] and UIUC lateral car database [9]

2 Feature Mining

In this section we construct a domain specific feature class by taking knowledge from the object class into account. A typical pool of features consists of up to several hundred thousand elements [1, 10–14]. Usually, the features have little resemblance to distinctive object characteristics. Our intention is to develop a more conspicuous feature which has a stronger correspondence to the object characteristics. In that way, the overall amount of features in the pool can be reduced by using a smaller set of more suitable 2Rec-Features. This enables the application of a more versatile feature, the 2Rec-Feature, whose pool of possible features would be otherwise hardly manageable in the exhaustive search during training of a boosting-based machine learning framework. 2Rec-Features are a generalized variant of Haar-like features that are not constrained to connected regions.

2.1 2Rec-Features

2Rec-Features acquire domain knowledge by analyzing the mean image of the positive image class. A mean image inevitably loses contrast and sharpness compared to single training images. Edges are unclear and the overall picture is very blurry. Figure 3(a) shows the mean image created by using about 8000 images from our positive face training set.

In a first step the mean image is segmented to derive distinctive regions. One of the most important conditions of the segmentation method is the ability to work on low-contrast images. To generate different kind of features, varied in size and position, the segmentation method also needs to be customizable. We analyzed K-Means Clustering [15], Watershed segmentation [16] and the Superpixel implementation of Greg Mori[1] [18] in order to find distinctive object characteristics. Neither K-Means nor Superpixel could cover content-related regions. The regions were too small or distorted by small scattered segments, as presented in Figure 3(b) and 3(c). In comparison Figure 3(d) shows a segmentation obtained from the Watershed algorithm that meets the requirements.

[1] The idea of Superpixels was originally developed by Xiaofeng Ren and Jitendra Malik [17].

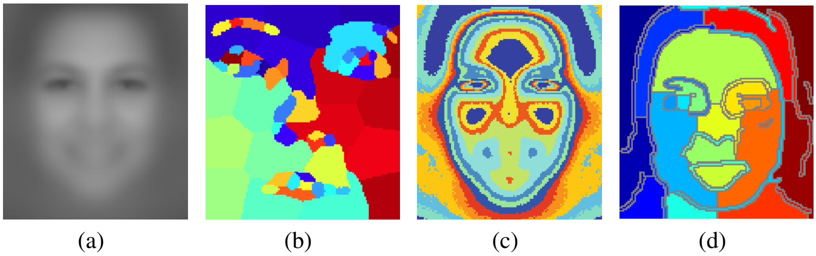


Fig. 3. (a): Mean face (b): Superpixel image (c): Clusters obtained with KMeans clustering (d): Watershed segmented regions

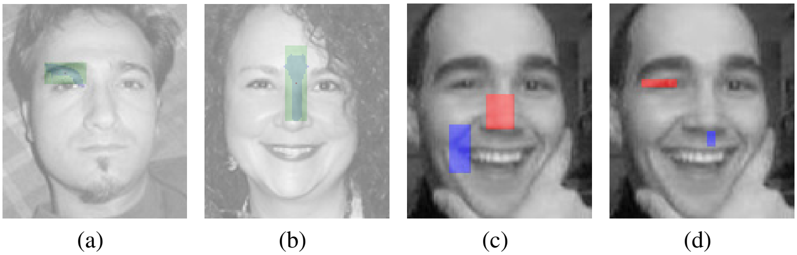


Fig. 4. (a) and (b): Segmented regions and approximated rectangles. (c) and (d): 2Rec-Features selected by the AdaBoost algorithm in the application of face detection.

The Watershed method was first formulated in 1978 by Digabel and Lantuéjoul [19, 20] followed by numerous improvements and modifications. Our Watershed method is based upon the algorithm by Fernand-Meyer [16]. The basic principle is described in a very intuitive way: The image is interpreted as a topographic relief and the gray level pixel are considered as altitude information. The idea is to flood the relief until the water reaches local minima. The boundary between different local minima (the watershed) is used to split different regions.

In a second step the segmented areas are approximated by rectangular regions to preserve the fast and efficient computation using the integral image representation. In order to create rectangles that approach the Watershed segments in terms of covered region we simply compute the mean and standard deviation of each segment and use these as position and size of the rectangles. Figure 4(a), 4(b) presents examples for Watershed segments and the derived approximated rectangles. A 2Rec-Feature is created by combining two of these spatially separated rectangles allowing to cover non-coherent regions. The feature value is composed by the subtraction of the pixel sums of both areas. These steps are repeated using different parameterizations of the segmentation method to generate a pool richer in features of varying size and position.

Overall, in the case of face detection 44732 distinctive 2Rec-Features get extracted as combinations of approximated rectangles provided by using varying parameters in the proposed segmentation method. Figure 4(c),4(d) shows exemplary a 2Rec-Feature. For the lateral car-detection 9200 features (see Figure 5) are generated.

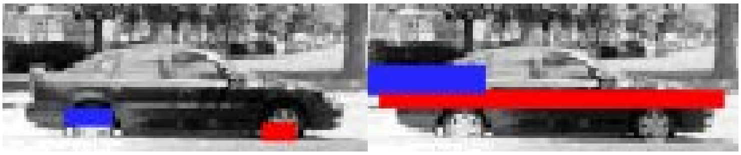


Fig. 5. 2Rec-Features selected by the AdaBoost algorithm in the application of lateral car detection

Moreover, our proposed feature mining approach is part of an initial offline pre-calculation step avoiding additional processing during training and detection.

3 Boosting Framework

The following section briefly introduces the machine learning algorithm AdaBoost, cascade training and margin analysis. This insight in the training procedure is required for the analysis of the training success in Section 4.1.

3.1 AdaBoost

Adaptive Boosting (AdaBoost) is a well-known machine learning algorithm introduced into object detection by Viola and Jones in 2001 [1]. The Viola and Jones framework applies AdaBoost to build a strong classifier from a set of Haar-like features:

Given a feature set and pairs (x_i, y_i) of training images x_i and labels y_i with $y_i = 0, 1$ for negative and positive examples respectively, a initial weight w_i is assigned to each training image with respect to the total number of negative and positive training images. In $t = 1, ..., T$ rounds, weak classifiers h_j are trained for each single feature that return 0 or 1 in case of a negative or positive classification, respectively. With regards to the classification error $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$, AdaBoost selects in each training round the weak classifier h_t having the lowest error ϵ_t and combines them to a strong classifier $H(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$. Where $\alpha_t = log(\frac{1-\epsilon_t}{\epsilon_t})$ is used to weight the weak classifier h_t and to update the image weights after each training round.

3.2 Cascade Training

To increase detection performance while reducing computation time, strong classifiers $H(x)$ are arranged in a cascade with increasing stage complexity. An image has to pass all stages in order to be classified positively. After training each stage, which consists of a single strong classifiers $H(x)$, the partial cascade is applied to the negative training set to delete correctly classified examples. The negative set is then replenished by a bootstrapping process that uses the partial cascade to collect false-positive examples from unseen images.

3.3 Margin Analysis

The training error of a boosted classifier is often not sufficient to analyze its generalization error. For this reason Freund and Shapire [21] proposed to take the confidence in classifications derived from the complete training set into account. A measure of this confidence is the margin that is defined as the difference between the sum of weights $\alpha_t = log(\frac{1-\epsilon_t}{\epsilon_t})$ of the weak classifiers h_t voting for the correct object class and the maximal sum of weights delegated to an incorrect class. The total sum of classifier weights is normalized to one so that the margin is a value in the range $[-1, 1]$. Hence a positive margin yields to a correct decision. Freund and Shapire pointed out that AdaBoost focuses on reducing the amount of examples having a small margin due to the update of the example weights in training. The distribution of margins on the training set is analyzed in Section 4.1 in order to evaluate the impact of different feature sets on the training success.

4 Experimental Results

In this section, classifiers based on 2Rec-Features and Haar-like features compliant with Viola and Jones are evaluated on different test sets and object classes. Results are presented for the MIT+CMU frontal face database [7], the Face Detection Data Set and Benchmark [8] and the UIUC lateral car database [9]. For comparison, results of other state-of-the-art methods are provided. Statistics of the training process are analyzed as well to show the quality of our proposed feature class.

4.1 Face Detection

The evaluated face detectors are trained on the ”MPLap GENKI-4K” database from the Machine Perception Laboratory in California [22] that consists of 4000 faces under different facial expressions. The training images are aligned to the eye positions and supplemented with mirrored images to a final training set of approximately 8000 images.

Training Success. In the first experiment, statistics of the training process are collected to analyze the training success of classifiers employing 2Rec-Features compared to conventional Haar-like features. Single-stage classifiers are trained to measure the progress of the margin distribution on the training set.

Figure 6(a) and 6(b) show the cumulative margin distribution of a 2Rec-Classifier and a Viola and Jones classifier after different rounds in training. In case of the 2Rec-Classifier the best feature is selected from a pool of 44732 features while the Haar-like feature pool supplies $4.61 \cdot 10^6$ features. Nevertheless, the training algorithm is able to select more discriminative features out of the smaller pool of 2Rec-Features. This can in particular be observed on the red margin curves in Figure 6. These curves present the margin distribution after two rounds in training using solely 2Rec-Features and Haar-like features, respectively. The values of the cumulative distributions at a margin of -1 shows that the amount of training examples that have been misclassified by both learned classifiers derived after two training rounds is significantly higher in case of Haar-like features. Similarly, both learned Haar-like classifiers classify approximately 52% of the training examples correctly, whereas both 2Rec-Classifier decide correctly for roughly 74% of the training set.

In other words, the more shallow progress for low margins of the 2Rec-Feature curves compared to the Haar-like feature curves after equal training rounds demonstrates that the 2Rec-classifiers are not only classifying a bigger part of the training set correctly but also that these decisions are more clearly. This advantage persists during training. After 100 training rounds the combined 2Rec-Classifier shows at the zero margin position, in contrast to the Haar-like classifier, no misclassifications anymore.

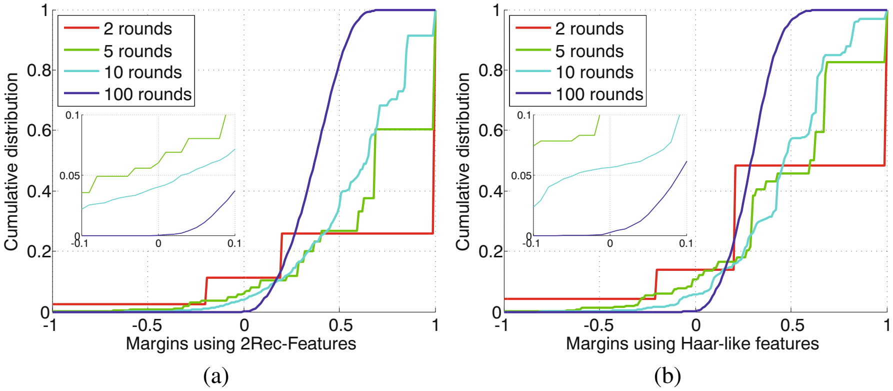
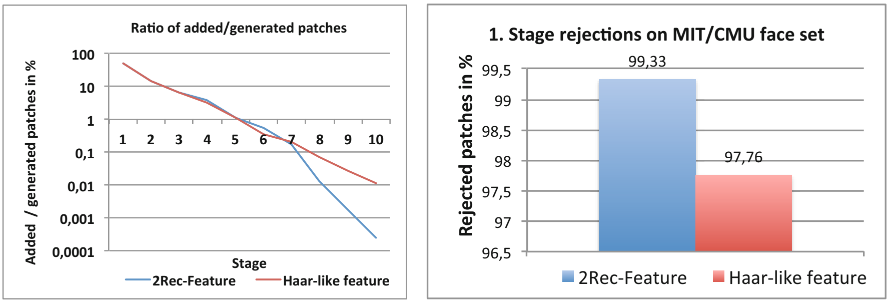


Fig. 6. Cumulative margin distributions with detail view of face training set after 2, 5, 10 and 100 rounds in case of (a): 2Rec-Features, (b): Haar-like features

A second experiment is conducted during the training phases of cascaded classifiers. The bootstrapping framework supplements the training framework with false-positives taken from a classification process on unseen negative images. The ratio between added and total generated sub-windows, in this case a false positive rate, can be used as a measure to assess the quality of a detector. Figure 7(a) shows this ratio over all trained stages. In the case of the 2Rec-Detector significantly more sub-windows have to be sampled to replenish the negative set.

Training Time. The benefit of our method in processing time in the feature selection of the training phase is obvious. Referring to the previously reported sizes of the feature pools, the Haar-like feature space is 103-times bigger than the 2Rec-Feature space. The feature selection process scales nearly optimal using feature pools of different sizes. Due to the constant computation overhead, we measure on a single multi-core workstation a slight reduction in efficiency per feature of 3.84% when processing the smaller pool of 2Rec-Features. This yields a total training speed-up of 99.04 using 2Rec-Features.



(a) (b)

Fig. 7. Face detection: (a) Diagram of ratio between added and generated patches at bootstrap framework, (b) Amount of rejected sub-windows within the first stage

Detection Performance. The first detection tests are performed on the MIT+CMU frontal face database [7] which consists of 130 grayscale images containing 511 faces. Despite its age, this test set is still challenging. The image database is partially noisy and blurred and contains several difficult samples like comics, line drawings and a binary raster image. In the evaluations on this test set we intentionally relinquish post-processing steps like non-maxima suppression. Hereby the results are focused on the influence of the different feature classes.

Figure 7(b) presents the rejection rates of the first stage of a 2Rec-Classifier and a Viola and Jones classifier measured on the complete MIT+CMU dataset. It can be observed that the first stage of the 2Rec-Classifier rejects more sub-windows than the comparable Viola and Jones classifier. This leads to a faster cascade performance.

The following experiment compares a 2Rec-Feature and a Haar-like feature classifier trained consuming equal training time. Due to the higher discriminative behavior and the faster learning performance, the 2Rec-Detector clearly outperforms the Viola and Jones detector, see Figure 8(a). The impact of this improvement can be noticed in Figure 8(b) and 8(c) that show detections in several MIT+CMU test images. The Viola and Jones detector produces significantly more false-positives.

Figure 9(a) shows the results of a 2Rec-Features detector on the MIT+CMU dataset compared to polygonal Haar-like features proposed by Pham et al. [13] and the extended Haar-like features of Lienhardt and Maydt [10]. The evaluation against the polygonal features is in particular interesting as these represent a recent development in improving Haar-like features. Pham et al. report feature pools in training of 210000 polygonal features and 86000 extended Haar-like features. The feature pool of the 2Rec-Features detector has a size of 44732. The detectors consist of a comparable number of features employing 200 polygonal features, 200 extended Haar-like features and 193 2Rec-Features. The results of the detectors based on polygonal and extended Haar-Like features are taken from [13]. The 2Rec-Features generate superior detection rates indicating that 2Rec-Features are able to describe more distinguishing object characteristics in spite of their smaller feature pool.

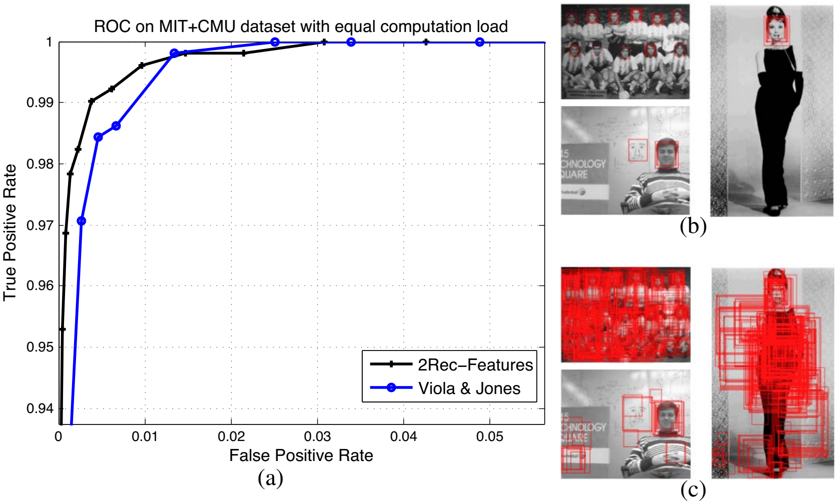


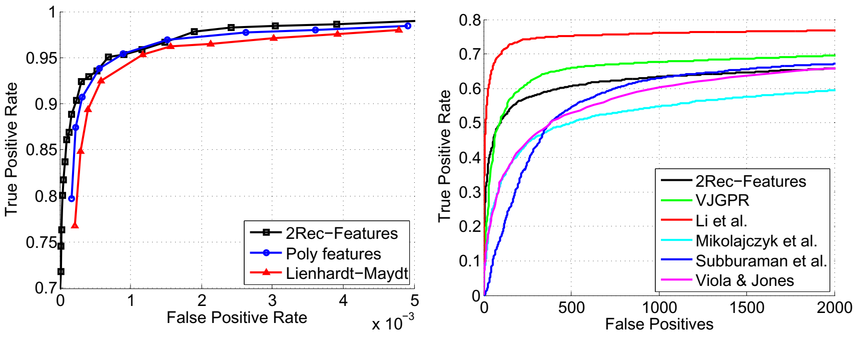
Fig. 8. Detections on the MIT+CMU database using detectors at similar computation loads. (a) ROC curve comparing 2Rec-Detector with Viola & Jones detector, (b) Example detections of 2Rec-Detector, (c) Example detections of Viola & Jones detector

To allow for a comparison of our proposed 2Rec-Detector to other state-of-the-art methods we conduct experiments on a newer test set, the Face Detection Data Set and Benchmark [8].

This test set consists of 2845 images containing 5171 faces and provides an evaluation tool to uniformly measure the quality of different methods. The evaluation procedure requires multiple detections to be merged to a single detection in advance, that is done in our method by mean shift clustering [23]. The authors supply results for different face detectors generated by this tool in form of ROC point files on the project web page[2].

Figure 9(b) presents the result of the 2Rec-Detector in comparison to several state-of-the-art methods. The 2Rec-Detector shows an overall competing performance and generates in particular for high precision very good results. Our proposed 2Rec-Features are a generalized variant of the conventional Haar-like features that allow to represent disjoined image regions but maintain the efficient computation. Hence, our main emphasis is to prove their advantage over Haar-like features that is clearly shown in Figure 9(b). Computational more complex methods like the results of Li et al. [24] and Jain et al. (VJGPR) [25] shown in Figure 9(b) use SURF features [24] or model inter-detection dependencies to exploit scene information [25] to derive higher detection rates. But especially Li et al. [24] propose besides SURF features a AUC score for cascade training as a second contribution that also adds to their good results and is not in contrast to 2Rec-Features. Adding such improvements to 2Rec-Features might increase their performance but also would obscure the impact of our main contribution.

[2] `http://vis-www.cs.umass.edu/fddb/results.html`



(a) (b)

Fig. 9. (a) ROC curves on MIT+CMU dataset using comparable number of different features. 200 features each for Poly features [13] and Lienhardt-Maydt [10] and 193 2Rec-Features. ROC line points are taken from [13]. (b) ROC curves on FDDB. Comparison of 2Rec-Features to different methods: VJGPR [25], Li et al. [24], Mikolajczyk et al. [26], Subburaman et al. [27] and Viola & Jones OpenCV implementation [28].

4.2 Lateral Car Detection

The lateral car database was collected by the Cognitive Computation Group of the University of Illinois [9]. It contains 1050 grayscale training images (550 cars and 500 non-car images). This database also provides test images and a evaluation tool to automatically calculate precision and recall. The multi-scale test set consists of 108 test images that contain 139 cars at different scales.

The performance of the 2Rec-Detector is compared to a Viola and Jones detector and to the connected-control-points detector proposed by Moutarde et al. [12]. Figure 10 presents the results in a precision-recall diagram. The comparison to connected-control-point features is in particular interesting as these features are also able to

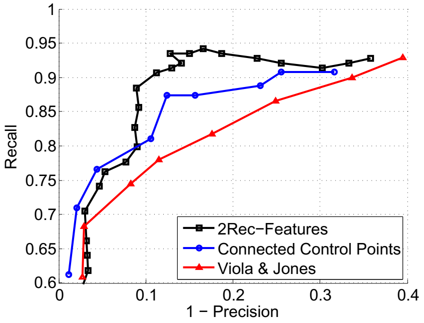


Fig. 10. ROC curve on UIUC multi-scale lateral car database. Comparison of 2Rec-Features to connected-control-points features [12] and Haar-like features [12]. The 2Rec-Detector shows competing up to superior performance. ROC line points of other methods are taken from [12].

describe spatially separated object characteristics and are used in a detector trained by the AdaBoost algorithm. But in contrast to 2Rec-Features the method of Moutarde et al. [12] exploits no domain knowledge in the creation of the feature pool. Hence, Moutarde et al. report a pool of $\sim 10^{19}$ features in training of the lateral car database. A typical Viola and Jones detector as the OpenCV [28] implementation also uses rotated Haar-like features and derives from its 14 basis features in case of the lateral car database (100×40 training images) a pool of $\sim 4.8 \cdot 10^6$ features in total. In comparison, the proposed 2Rec-Feature detector uses only 9200 features in training.

Despite its much smaller feature pool the 2Rec-Detector achieves competing up to superior results.

5 Conclusion

This paper proposed a feature mining algorithm that constructs a feature pool for boosting-based object detectors. In this process, domain knowledge provided by the positive training set is extracted and incorporated into the proposed 2Rec-Feature class. 2Rec-Features preserve the benefit of efficient computation using integral images. But in contrast to conventional Haar-like features, spatially separated regions can be represented that describe distinctive object characteristics. Furthermore, the size of the feature pool is drastically decreased as the portion of redundant and irrelevant information is reduced. Evaluations of the training process demonstrate a higher training success using the much smaller pool of 2Rec-Features compared to Haar-like features indicating the higher discriminative power of our proposed feature class. Experiments were conducted on the MIT+CMU frontal face database, a recent database for unconstrained face detection (FDDB) and the UIUC lateral car database. Detectors trained on 2Rec-Features showed improved detection performance and superior precision compared to conventional Haar-like features and state-of-the-art methods.

References

1. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
2. Dollár, P., Tu, Z., Tao, H., Belongie, S.: Feature mining for image classification. In: CVPR. IEEE Computer Society (2007)
3. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. In: Artificial Intelligence, vol. 97, pp. 245–271 (1997)
4. Koller, D., Sahami, M.: Toward optimal feature selection. In: ICML, pp. 284–292. Morgan Kaufmann (1996)
5. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182 (2003)
6. Zhao, X., Chai, X., Niu, Z., Heng, C., Shan, S.: Context constrained facial landmark localization based on discontinuous haar-like feature. In: FG, pp. 673–678. IEEE (2011)
7. Sung, K.K., Poggio, T., Rowley, H.A., Baluja, S., Kanade, T.: MIT+CMU frontal face dataset a, b and c. MIT+CMU (1998)
8. Jain, V., Learned-Miller, E.: Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst (2010)

9. Agarwal, S., Awan, A., Roth, D.: UIUC image database for car detection (2002)
10. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. IEEE ICIP 2002 (2002)
11. Stanciulescu, B., Breheret, A., Moutarde, F.: Introducing new adaboost features for real-time vehicle detection. In: COGIS (2007)
12. Moutarde, F., Stanciulescu, B., Breheret, A.: Real-time visual detection of vehicles and pedestrians with new efficient adaboost features. In: IEEE IROS (2008)
13. Pham, M.T., Gao, Y., Hoang, V.D.D., Cham, T.J.: Fast polygonal integration and its application in extending haar-like features to improve object detection. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010) (2010)
14. Smith, K., Carleton, A., Lepetit, V.: Fast ray features for learning irregular shapes. In: IEEE: ICCV, pp. 397–404 (2009)
15. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society 28(1), 100–108 (1979)
16. Meyer, F.: Topographic distance and watershed lines. In: Signal Processing, vol. 38, pp. 113–125 (1994)
17. Ren, X., Malik, J.: Learning a classification model for segmentation. In: 9th Int. Conf. Computer Vision, vol. 1, pp. 10–17 (2003)
18. Mori, G., Ren, X., Efros, A., Malik, J.: Recovering human body configurations: Combining segmentation and recognition. In: IEEE Computer Vision and Pattern Recognition (2004)
19. Lantuéjoul, C.: La squelettisation et son application aux mesures topologiques des mosaiques poly- cristallines. PhD thesis, Ecole des Mines (1978)
20. Digabel, H., Lantuéjoul, C.: Iterative algorithms. In: Actes du Second Symposium Européen d'Analyse Quantitative des Microstructures en Sciences des Matériaux, pp. 85–99 (1977)
21. Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.S.: Boosting the margin: A new explanation for the effectiveness of voting methods. In: Proceedings of the Fourteenth International Conference on Machine Learning (ICML), pp. 322–330 (1997)
22. The MPLab GENKI Database, u.S.: `http://mplab.ucsd.edu`. - (-)
23. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 603–619 (2002)
24. Li, J., Wang, T., Zhang, Y.: Face detection using surf cascade. In: ICCV Workshops, pp. 2183–2190. IEEE (2011)
25. Jain, V., Learned-Miller, E.G.: Online domain adaptation of a pre-trained cascade of classifiers. In: CVPR, pp. 577–584. IEEE (2011)
26. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 69–82. Springer, Heidelberg (2004)
27. Subburaman, V.B., Marcel, S.: Fast Bounding Box Estimation based Face Detection. In: ECCV, Workshop on Face Detection: Where We Are, and What Next? (2010)
28. Bradski, G.: The opencv library. Dr. Dobb's Journal of Software Tools (2000)