International Journal of Computer Vision and Signal Processing, 5(1), 14-22(2015) ORIGINAL ARTICLE

Enhanced Contour Description for Pedestrian Detection

Xiaoyun Du, Robert Laganière*, Si Wu

Shanghai Research Institute of China Post Group,
Shanghai 200062, P.R. China.

VIVA research lab, School of Electrical Engineering and Computer Science
University of Ottawa, ON, K1N 6N5, Canada.

School of Computer Science and Engineering, South China University of Technology,
Guangzhou 510006, P. R. China.

Abstract

In this paper, an enhanced contour descriptor is presented for pedestrian detection, an important and challenging task in computer vision. A discriminative feature is essential to build a robust detection. Histograms of Oriented Gradients (HOG) and Local Binary Patterns (LBP) are well known descriptors with proven efficiency. We propose here the use of Variational LBP (VLBP) as an auxiliary feature and combine it with HOG to generate a discriminative contour descriptor used in our detection model. We also perform feature selection by using the Feature Generating Machine in order to select the most useful elements of a descriptor vector without impacting on the classification performance. Moreover, a two-layer cascade model is proposed to achieve both accurate detection and lower computational complexity. The intersection kernel based support vector machine is employed in our cascade model as a performing classification tool that integrates well histogrammic features. A bootstrapping algorithm is also applied in our training procedure to improve the performance of classification. The result of the experiment shows that our approach achieves good performances on the INRIA benchmark dataset. A conclusion of this work is that an enhanced and concise descriptor that combines HOG and VLBP can improve contour description and thus leads to better detection performance.

Keywords: Pedestrian detection, contour description, feature selection, cascade, bootstrapping

ISSN: 2186-1390
http://www.cennser.org/IJCVSP

Article History:
Received: 10 February 2015
Revised: 29 April 2015
Accepted: 29 May 2015
Published Online: 31 May 2015

1. INTRODUCTION

Pedestrian detection is an active area of research in computer vision. This topic draws attention because of its usefulness in many applications, such as intelligent surveillance, driver assistance system, and event detection. Moreover pedestrian detection is also considered as the fundamental component for several high-level systems, such as pedestrian tracking, people re-identification and action recognition. Pedestrian detection remains a challenging matter because of the complexity of the background environment, the variability in shapes of pedestrians' body and the wide diversity in pedestrian's clothing and postures.

The identification of a discriminative feature is essential to the design of a robust detection system. A wide variety of descriptors has been proposed in the recent literature, several of them are presented and analyzed in [1]. Among them, the Histograms of Oriented Gradients (HOG) [2] [3] [4] has been shown to be one of the best performing features, and it has been used in several existing methods. HOG extracts contour information and is invariant to geometric and photometric transformations which makes it particularly suited for pedestrian detection. Another notable descriptor is the Local Binary Patterns (LBP) [5] that extracts the textural information of an image patch. It has been successfully applied to face detection [6] and object segmentation [7]. Some works have been proposed that combine HOG and LBP for improving pedestrian detection

*Corresponding author
Email addresses: duxysh@gmail.com (Xiaoyun Du), duxysh@gmail.com (Si Wu), laganier@eecs.uottawa.ca (Si Wu), ezwusi@gmail.com (Si Wu)

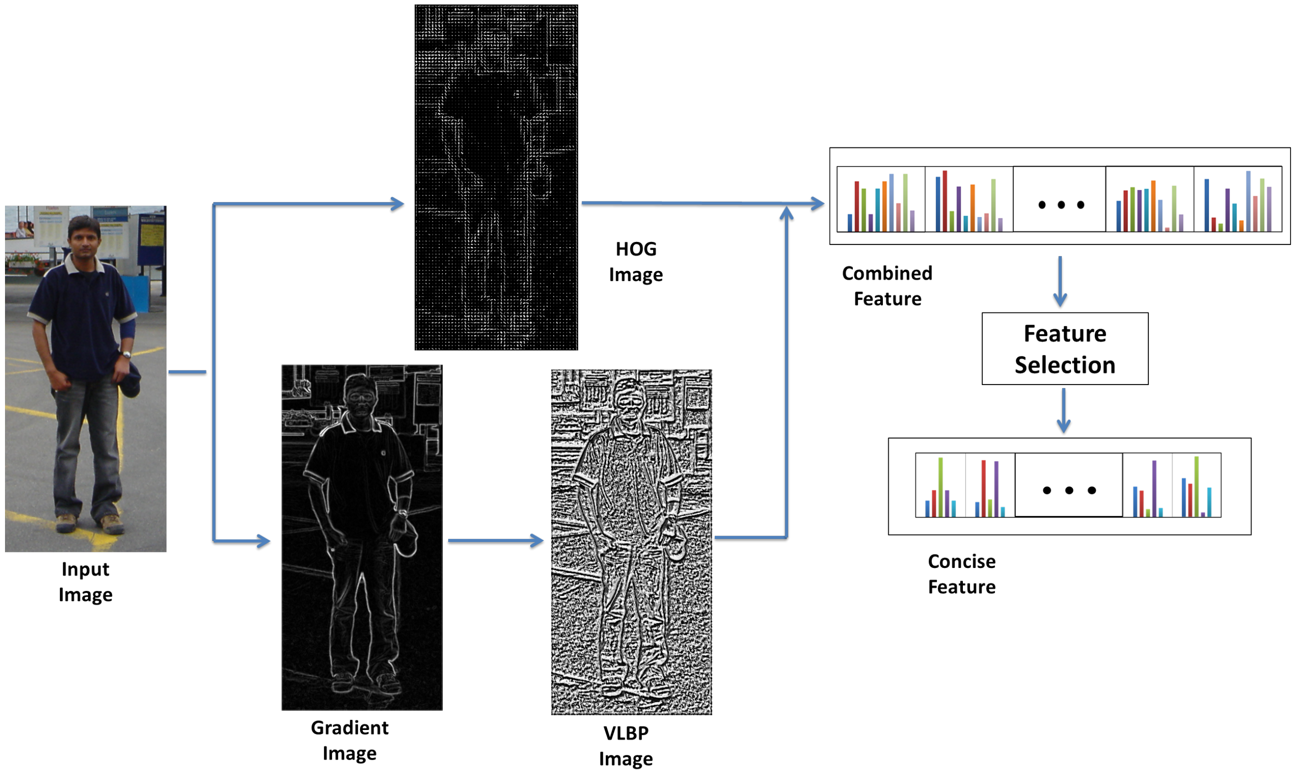

Figure 1: The procedure of our enhanced and concise descriptor.

[8]. However, LBP is designed to describe texture information rather than contour information. For the problem of pedestrian detection, contour information usually plays a more important role than texture. Indeed, pedestrian wear clothes showing a wide diversity of patterns and colors and, as a result, no specific textural pattern can be considered as a good indicator of the presence of a person in an image. Variational LBP (VLBP) has been proposed by Wu et al. in [9] [10] to extract contour information in the context of scene categorization and object detection. VLBP is based on the image gradient, which focuses more on contour information while disregarding contextual details, conveyed by the original image. In order to provide a better characterization of the contour information in an image patch, we combine here HOG and multi-scale VLBP and propose an enhanced descriptor to improve detection performance. But the merging of these two descriptors result in a high-dimensional representation; this one can then be made more concise by applying feature selection. Usually Principal Components Analysis (PCA) [11] [4] and Partial Least Square (PLS) [12] are popular approaches to achieve this goal. In this work, we use the Feature Generating Machine (FGM) [13] to generate a concise descriptor which preserves the most important parts of the combined feature without significantly impacting the classification performance. Figure 1 shows an overview of the computational process for the feature descriptor introduced in this paper.

To build a powerful detection system, we also propose a two-layer cascade classifier as shown in Figure 2. The first layer of this classifier uses a classic linear SVM for fast classification while a Histogram Intersection Kernel based Support Vector Machine (HIKSVM) [14] is used in the second layer to improve the accuracy of the classification. In addition, our training procedure has recourse to bootstrapping [15] [16] [2] [17] [18], in order to improve the classification performance with a limited number of training samples.

The main contribution of this work is an enhanced contour descriptor that combines HOG and VLBP in an efficient two-layer cascade classifier. We show that this model outperforms the baseline detection model of HOG linear SVM while keeping the computational complexity low. This paper is organized as follows: In Section 2 a brief referral of the related works is presented. In Section 3 we describe the contour descriptor used in our work. Section 4 explains our 2-layer cascade detection framework and the techniques used in the feature selection, classification and training procedures. In Section 5 we show the result of experiments on the INRIA benchmark dataset. Finally in Section 6 a conclusion of this paper is made.

2. RELATED WORK

Pedestrian detection has been the object of several works in the recent literature [1] [19]. In an effort to improve the detection performance, these works have been focusing on

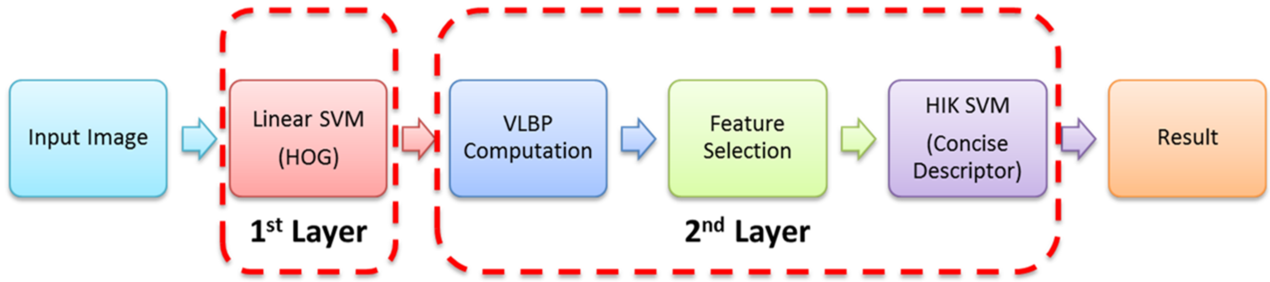

Figure 2: General framework of the proposed detection system

two main aspects: feature description and classification algorithms.

The Histograms of Oriented Gradient (HOG) descriptor is a milestone feature [20]. This descriptor conveys the contour information of objects in images. The fundamental aspect of this descriptor is that a robust and invariant representation of a target object can be obtained from the distribution of its intensity gradient orientations. Specifically, a detection window is divided into small connected regions (cells), and for each cell, a histogram of edge orientations is calculated. The gradient histograms from these different cells are then concatenated together in order to produce a single descriptor. HOG shares some similarities with the Scale Invariant Feature Transformation (SIFT) [21]. HOG is calculated on a dense array of overlapped sample windows while SIFT uses sparsely distributed descriptors based on the interest points. SIFT is often used to identify object categories using a bag of words framework[22]. Zhu et al. proposed fast calculation of HOG descriptor in [23] by using integral histograms.

Texture information can also be used to describe the characteristics of an object. To describe texture information, the LBP descriptor was first proposed in [5]. It is nowadays a classic texture descriptors and has been used in several areas, for instance, face recognition [6], face detection [24], [25], and object segmentation [7]. This descriptor is invariant to change in illumination, which makes it to perform well and robustly in classification and segmentation task. Other types of LBP descriptors, have been proposed to improve its performance and to meet specific requirements, such as Elongated Local Binary Patterns (ELBP) [26], overlapped LBP (oLBP) [27], Rotation Invariant LBP [28]. Of particular interest is the variational LBP descriptor that have been proposed by Wu, et al. in [9]. This variational LBP (VLBP) descriptor is used to extract the contour and texture feature from the gradient image.

Color information can also be a useful cue for improving pedestrian detection accuracy. The Color Self Similarity (CSS) descriptor was introduced by Walk et al. in [16]. The author proposed a new feature descriptor that uses color information with the idea that color patterns of pedestrian's clothes can potentially offer additional information that can be exploited by a classifier. This color self similarity feature is generally computed in HSV color space. The pairwise similarities of the features' histograms are computed and histogram intersection is used as the distance function. However by using co-occurrence histograms, a second order image statistics, the dimension of the feature descriptor becomes high. Note also that this descriptor is usually employed as a complement to other features such as the HOG.

As an another important system component, an effective algorithm of classification is also essential to a powerful detection system. Support Vector machines (SVM) is a leading technique used in vision tasks such as pedestrian detection [2], [29] , face detection [30], and car detection [31]. SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, and is mostly used for classification and regression analysis.

The deformable part models (DPM) [32] [15] [33] [34] [35] is probably the most successful approach for object detection based on HOG. It achieves state-of-the-art people detection [36] through the use of a latent SVM formulation. This DPM approach is trained using a discriminative procedure that only requires the bounding boxes of the objects, leading to an efficient and accurate detector. By using this elegant framework, DPM can capture variations in appearance. An important benefit of this approach resides in the fact that a part-based model is trained without an explicit labeling of the part locations. Part-based models are generally better at handling pose variations in the object model. However, they are computationally more complex even if efficient detection mechanism can be introduced such as the star-cascade presented in [37].

An approach based on feature pooling is proposed in [38]. In this method, a spatial pooling operator over low-level features is introduced in order to improve the recognition rate. Classifier learning is based on AdaBoost in which the partial area under the ROC curve (pAUC) is optimized. They observe that by combining multiple features, detection performances are generally improved. Recently, good recognition performances have been obtained using deep networks. In [39], a deep learning framework that learn features jointly with deformation and occlusion handling components. A deep model that automatically learns multi-scale scene-specific features in static video surveillance is proposed in [40]. In this approach, the target training samples are automatically selected and labeled. An objective function is introduced that encodes the confidence scores of each training sample. This way, the deep model becomes robust to labeling mistakes on target training samples.

In this work, our objective was to design an enhanced pedestrian detector mainly by improving the contour description. We found that when HOG is combined with VLBP, the enhanced descriptor is better than either HOG, LBP or VLBP. The computational constraint is a major requirement in applications such as pedestrian detection for driver-assistance systems in which the detection algorithm needs to be implemented on a low-power hardware architecture having severe memory limitations and fixed-point constraints. When feature selection is employed, our enhanced descriptor is concise without impacting on the performance.

3. ENHANCED CONTOUR DESCRIPTION

This section briefly presents the HOG descriptor and the Variational LBP descriptor for describing contours information of pedestrian in images. We use VLBP as an auxiliary feature and combine it with HOG to generate our discriminative descriptor.

3.1. HOG

The HOG descriptor is well-known as a robust contour descriptor and has been widely used in many applications. The main idea of this feature is that contours of objects in an image can be described by using the distribution of intensity gradient directions. To compute the HOG descriptor, a detection window is divided into non-overlapped 8×8 cells. For each cell, a histogram of gradient orientations made of 9 bins is computed and weighted by gradient magnitudes. Each 2×2 neighboring cells constitute a block, and the corresponding 4 histograms are concatenated to a 36-dimensional feature vector followed by a L2-normalization. For a 64×128 detection window, there is a total of 105 overlapping blocks when the block stride is set to 8×8 pixels. The corresponding HOG descriptor is built by concatenating all features from these 105 blocks. As a result, the HOG descriptor has 3780 dimensions.

3.2. VLBP

The VLBP descriptor , which was proposed by Wu et al.[9], is a variant of LBP, that extracts edge information in a detection window. Instead of using the original gray image, the computed gradient magnitude image is used to produce the corresponding LBP image. For each scanned window, the VLBP descriptor is composed of the histogram features of the blocks in the LBP image. Specifically, the gradient G is computed by using a pair of Sobel filters given an input image I. Let $M = |G|$ denotes the magnitude of G. The LBP image L is computed on M, at pixel p using the clockwise sequence of pixels q_i in the 3×3 neighborhood of p

$$
\begin{aligned}
Lp &= \sum_{i=0}^{7} 2^i \bullet s(p, q_i), \\
s(p, q_i) &= \begin{cases} 1, if\ p \geq q_i; \\ 0, otherwise; \end{cases}
\end{aligned}
\tag{1}
$$

The VLBP descriptor is made of the concatenated histograms computed over overlapping blocks inside the detection window. Finally the VLBP descriptor is normalized by L2-norm.

For a 64×128 detection window, we clip the center 48×96 region for VLBP feature extraction because we experimentally found that the border region of image in the INRIA training dataset fails to provide discriminative information. Similar to the HOG descriptor computation, a histogram of LBP values with 64 bins is computed in each 16×16 block. We use a block stride of 8×8 to obtain a total of 55 blocks, and concatenate the corresponding histograms to build a 3520-dimensional VLBP descriptor.

To capture more discriminative information, a multi-scale representation is usually adopted. In this work, we use three window sizes: 12×24, 24×48, 48×96. In the first case, the block sizes and strides are changed to 12×12 and 6×6 respectively. The proposed multi-scale VLBP descriptor is illustrated in Figure 3.

3.3. Enhanced Contour Descriptor

An enhanced contour descriptor is obtained here by combining the HOG and the VLBP descriptors, but the dimension of this combined descriptor vector is 8132 (HOG dimension 3780 + three-scale VLBP dimension 4352). This imposes a significant computational load on the classifier. We therefore introduce feature selection to simplify our combined descriptor.

In order to identify the most discriminative features without impacting on the classification performance, the Feature Generating Machine (FGM) [13] algorithm is applied, which is a very useful tool to perform feature selection on high dimensional data. Alternatively, for the purpose of dimensionality reduction, the PCA algorithm is often used. It is an unsupervised approach that uses an orthogonal transformation to projects high-dimensional data into a lower dimensional space. In both the training and test phases, the full feature vector needs to be computed before the projection. Different from PCA, FGM is a supervised method for feature selection. Specifically, FGM employs a sparse support vector machine to determine a subset of feature components for discriminant, while avoiding impacting on the classification accuracy. The selected components are considered to be more discriminative than

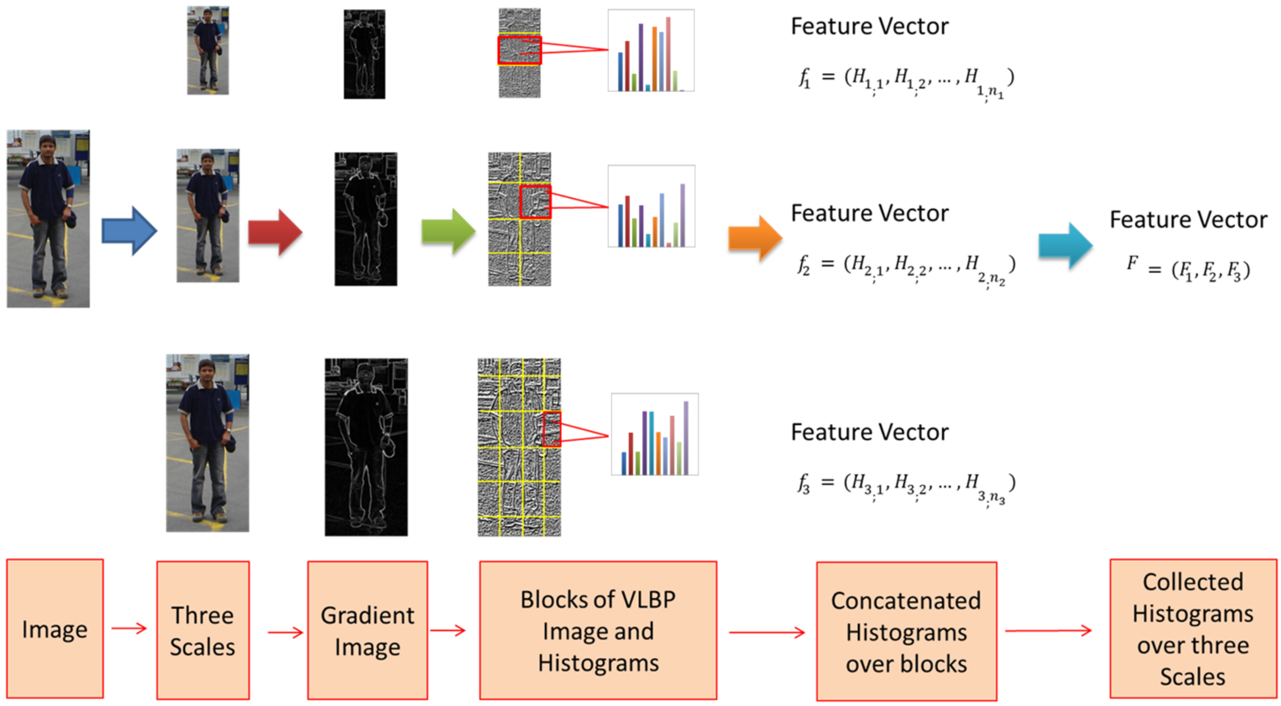

Figure 3: An example of multi-scale VLBP computation

the others. FGM feature selection is performed by repeatedly generating a pool of informative feature subsets. Specifically, the problem of feature selection is formulated as a SVM problem, in which the features are selected or rejected according to a 0-1 vector. After convex relaxation, the problem can be transformed into a convex multiple kernel learning problem and solved by an efficient and scalable cutting plane algorithm. It should be noted that FGM is performed in the training phase. After feature selection, only the selected feature components need to be computed such that less computation is required in the test phase.

The result of FGM provides an indication of the most informative features. In our case, the size of our enhanced contour descriptor is thus reduced by about 50%.

4. PROPOSED APPROACH AND IMPLEMENTATION

In this paper, a two-layer cascade model is proposed to get precise detection result while limiting the complexity of the procedure. The first layer is defined as a linear SVM classifier based on the HOG feature. It is used to generate candidates for the second layer. So we set a lower threshold than usual to make this layer accepting almost all positive samples while rejecting most negative samples. This way, only a few candidates need to be classified by the next layer. In the second layer, the HOG and the Variational LBP features selected from FGM are used to generate a more discriminative descriptor. Moreover, HIKSVM is employed as a powerful tool for improved classification. Bootstrapping algorithm is used during the training process in order to obtain a classifier with optimal performances. The following subsections provide more details on the training procedure.

4.1. Classification

Histogram intersection kernel combined with SVM is employed in this paper for effective classification. Many works have shown that HIKSVM has outstanding performance [14] [41] [42]. In addition, because HOG and our variational LBP are both histogram features, they are very suitable for being integrated into a HIKSVM framework. However, the complexity of intersection operation that HIK requires is proportional to the number of support vector. To solve this problem, Maji, et al. [14] proposed a fast implementation of the HIKSVM. Specifically, for an input vector X, the decision function of HIKSVM can be rewritten as follows:

$$
\begin{aligned}
h(X) &= \sum_{i=1}^{n} a_i y_i k(X, X_i) + c \\
&= \sum_{i=1}^{n} a_i y_i \left(\sum_{j=1}^{m} min(x(j), x_i(j)) \right) + c \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_i y_i \; min(x(j), x_i(j)) \right) + c \\
&= \sum_{j=1}^{m} h_j(x(j)) + c,
\end{aligned} \tag{2}
$$

where X_i and y_i are a support vector and its corresponding label respectively, a_i is the coefficient, c is a constant, and $k(x, y) = \sum_{i=1}^{n} min\,(x(i), y(i))$ is an intersection kernel. Based on Eq.(2), h_j can be cheaply computed by using a lookup table. We employ this efficient HIKSVM on the second layer of our framework.

4.2. Training Procedure

In the training procedure of our detection system, we use a bootstrapping algorithm, which was first proposed in statistics and consists in oversampling the sample data and applying some results back to the training data. In our case we use an initially trained classifier to re-scan the testing samples and randomly select the falsely detected results to update the training dataset. We then iteratively update the training dataset, and make it more relevant for training.

Given the pre-trained detector of HOG+SVM(linear) as the first layer, the initial training data for the second layer is generated from this first layer which includes all positive samples and some randomly selected negative samples. The HIKSVM is trained by using the HOG_VLBP feature on this initial training set. Next, a test is done on a testing subset, as described in the next section, to extract hard negative samples and used them to update our negative training data. After that, we start over the above training process based on the updated training dataset. There are two criteria for stopping this process. The first one is that the false positive rate on a restricted test set should be closed to 0 (in our cases, it should be less than 0.0001). The second one is that the total number of hard negative samples generated from the test should be smaller than a preset threshold (in our case, it should be less than 30). The two stopping criteria for the bootstrapping algorithm are set to make sure that the recognition performances have stabilized and that no more hard samples can be generated. Bootstrapping, which exploits the hard examples in the training procedure, is a process that does contribute much to the training procedure. The bootstrapping procedure is illustrated in Figure 4.

5. EXPERIMENTS

In this section, we evaluate our enhanced contour description for pedestrian detection on the INRIA dataset. The performance of our model is presented in form of a trade-off curve: miss rate ($1 - Recall$) (Y axis) and false positive per image (FPPI) (X axis). Given a specific FPPI, an approach achieving lower miss rate is better.

5.1. Dataset

The INRIA dataset is a classic and popular dataset for pedestrian detection. This dataset contains a large number of annotated upright pedestrians in still images which are taken from various viewpoints and with a variety of scenes captured under different lighting conditions. This dataset is split into two subsets: the training set and the testing set. There are 1208 annotated pedestrians , which are usually flipped horizontally to get 2416 images, in 614 full size positive image set, and 1218 full size negative image samples. Thus the initial dataset is composed of these positive samples (2416) and negative samples (1218). In addition there are 1126 annotated pedestrians in 288 images positive set and 453 negative image samples in the testing dataset. For our testing procedure during bootstrapping, we use a testing subset, which contains 453 original images, of the INRIA dataset. We then randomly extract 10 small 64×128 sub-images from each original image. Thus the test dataset for extracting hard negative sample contains 4530 non-pedestrian images.

5.2. Bootstrapping

We trained the two classifiers in our two-layer model through bootstrapping. In this experiment, we focus on the effectiveness of bootstrapping on the second layer in Figure 5. The curve of the initial round represents the performance of the model trained using all 2416 positive samples and 1126 initial random selected negative samples. The other curves are the results of the subsequent bootstrapping rounds retrained by the same positive samples and the updated negative samples, which consist of the samples of the previous round and new hard negative samples generated via the model scanning on the non-pedestrian testing images. For example, there are 2372 negative samples obtained after the first round, which is composed of 1126 initial negative samples and 1246 hard negative examples extracted by the trained model. In the experiment, the bootstrapping procedure is performed over 7 rounds. The number of negative samples obtained on the 7th round is almost the same as that on the 6th round, which means that only a very small number of hard negative samples are extracted and the 7th round improves only marginally. It can be observed that the performances of these 7 rounds are getting progressively better, which demonstrates that the additional hard negative samples generated through the bootstrapping algorithm contribute significantly to the resulting model performance.

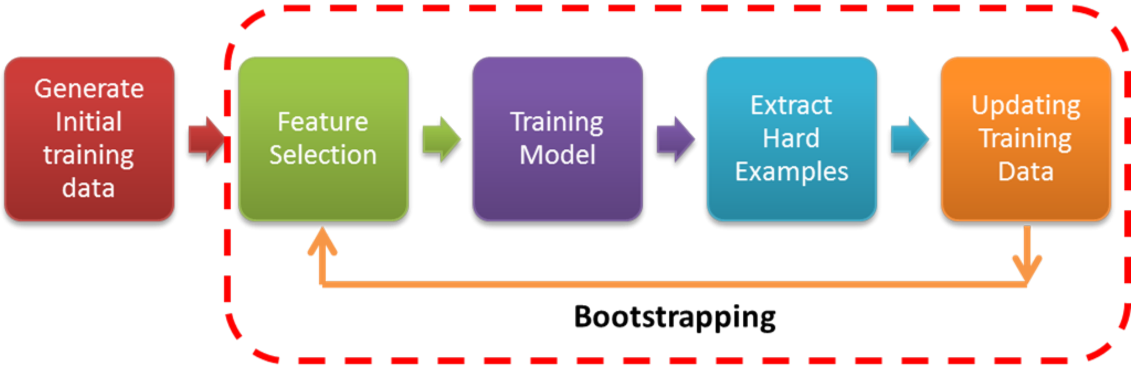

Figure 4: The procedure of bootstrapping.

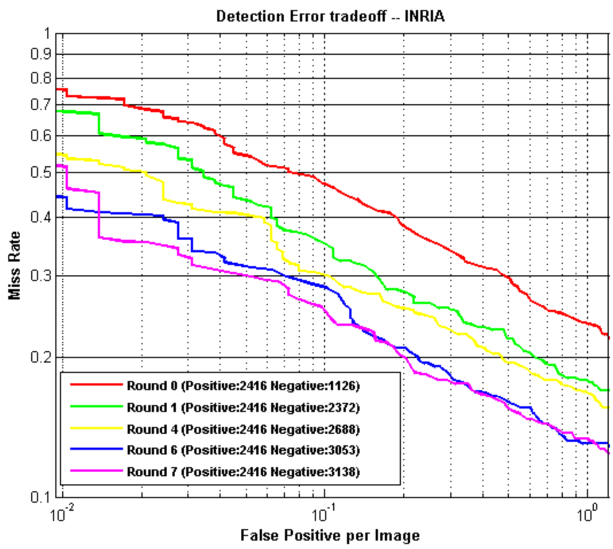

Figure 5: Performance of the proposed method after bootstrapping.

5.3. Comparison with contour descriptors

In order to validate our approach, we compare our approach with the following detectors: HOG detector, VLBP detector, HOG_LBP detector, HOG_VLBP detector, HOG_VLBP (Multi-Scale) detector (our proposed model). All these detectors are tested on the INRIA testing dataset. From Figure 6, we can see that the performances of VLBP are only marginally better than HOG. When we combine HOG and VLBP descriptors, the results are improved. This demonstrates that HOG and VLBP features are complementary to each other. The result of the HOG feature combined with the VLBP feature is also better than the HOG feature combined with the LBP feature, which shows that the variational LBP is better than the original LBP in our model. We used the multi-scale VLBP features on the second layer of our model, and the performance of the curve of the multi-scale VLBP feature is better than that of a single scale VLBP feature. Furthermore by using the FGM algorithm, we reduced the dimension of our enhanced descriptor by almost 50% (from 8132 to 4000 dimensions) without significantly impacting the performance. The detection speed of our proposed cascade model achieve 0.5 frame per second on the INRIA dataset on average. Some comparison results are shown in Figure 7.

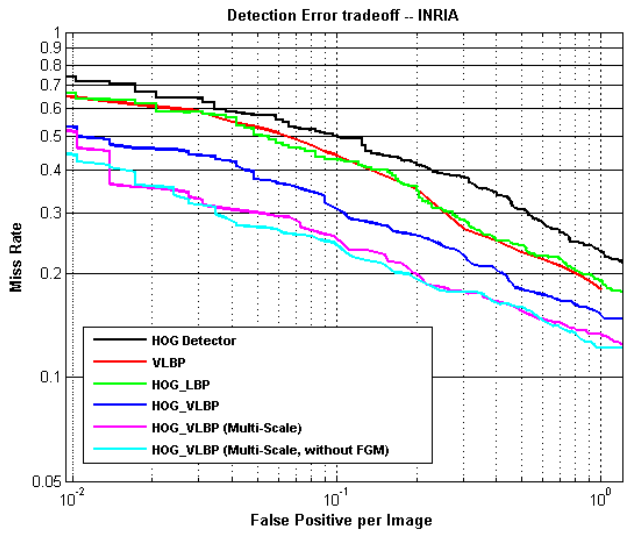

Figure 6: Comparison of contour descriptors.

5.4. Comparison with other approaches

We also compare our proposed method with other classic detector: The Viola-Jones detector [43] , DPM detector [32], and ConvNet detector [44]. From the Figure 8, we can find that our model is better than the Viola-Jones detector, which is a widely used detection system. Besides our approach is also better than ConvNet in some cases (when FPPI <0.1). When compared to the DPM method, considered as one of the state-of-the-art approaches, our method obtains competitive results at low FPPI, which is the usual operational requirement.

6. CONCLUSION

In this paper, we mainly focus on presenting an enhanced descriptor, which is combined with HOG and VLBP,

(a) (b)

Figure 7: Comparison of (a) HOG baseline method and (b) our enhanced contour detection method.

to better describe the contour information of pedestrian in images. Feature Generating Machine was used to generate a concise descriptor without impacting on the performance. Moreover we proposed a two-layer cascade to cooperate with this discriminative descriptor for pedestrian detection. Given a trained HOG+SVM(linear) detector on

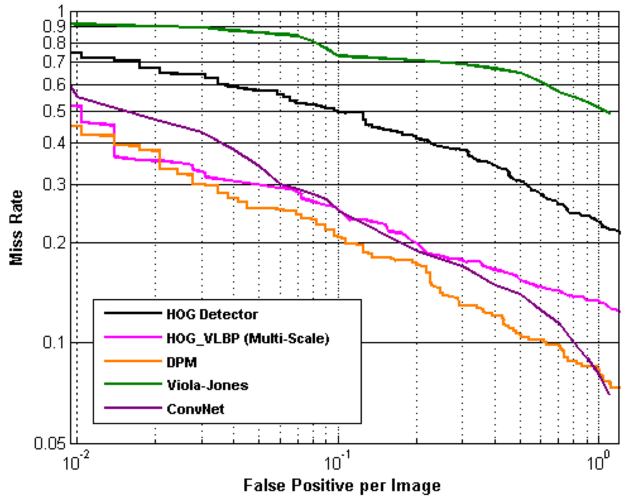

Figure 8: Comparison of our proposed method with other methods.

the first layer to quickly generate candidates, we trained an HIKSVM based detector to make a final decision in the second layer. As a powerful tool for classification, HIKSVM can be used in an efficient way through a fast implementation. Bootstrapping was used to produce an optimally trained classifier. The experiment results show that our enhanced contour model for pedestrian detection outperforms the baseline model on the INRIA dataset.

References

[1] P. Dollar, C. Wojek, B. Schiele, P. Perona, Pedestrian detection: An evaluation of the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (4) (2012) 743–761.

[2] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, 2005, pp. 886–893.

[3] F. Suard, A. Rakotomamonjy, A. Bensrhair, A. Broggi, Pedestrian detection using infrared images and histograms of oriented gradients, in: Intelligent Vehicles Symposium, IEEE, 2006, pp. 206–212.

[4] T. Kobayashi, A. Hidaka, T. Kurita, Selection of histograms of oriented gradients features for pedestrian detection, in: Neural Information Processing, Springer, 2008, pp. 598–607.

[5] T. Ojala, M. Pietikainen, D. Harwood, Performance evaluation of texture measures with classification based on kullback discrimination of distributions, in: Proceedings of the 12th International Conference on Pattern Recognition, Vol. 1, IEEE, 1994, pp. 582–585.

[6] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: European Conference on Computer Vision, Springer, 2004, pp. 469–481.

[7] M. Heikkilä, M. Pietikäinen, J. Heikkilä, A texture-based method for detecting moving objects, in: 15th British Machine Vision Conference, Vol. 1, 2004, pp. 187–196.

[8] X. Wang, T. X. Han, S. Yan, An hog-lbp human detector with partial occlusion handling, in: IEEE 12th International Conference on Computer Vision, 2009, 2009, pp. 32–39.

[9] J. Wu, J. M. Rehg, Centrist: A visual descriptor for scene categorization, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8) (2011) 1489–1501.

[10] C. G. Jianxin Wu, Nini Liu, J. M. Rehg, C4: A real-time object detection framework, IEEE Transactions on Image Processing 22 (10) (2013) 4096–4107.

[11] I. Jolliffe, Principal component analysis, Encyclopedia of Statistics in Behavioral Science, 2005.

[12] W. R. Schwartz, A. Kembhavi, D. Harwood, L. S. Davis, Human detection using partial least squares analysis, in: IEEE 12th International Conference on Computer vision, 2009, pp. 24–31.

[13] M. Tan, L. Wang, I. W. Tsang, Learning sparse svm for feature selection on very high dimensional datasets, in: Proceedings of the 27th International Conference on Machine Learning, 2010, pp. 1047–1054.

[14] S. Maji, A. C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[15] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (9) (2010) 1627–1645.

[16] S. Walk, N. Majer, K. Schindler, B. Schiele, New features and insights for pedestrian detection, in: IEEE conference on Computer Vision and Pattern Recognition, 2010, pp. 1030–1037.

[17] H. A. Rowley, S. Baluja, T. Kanade, et al., Human face detection in visual scenes, in: Technical Report CMU-CS-95-158, School of Computer Science, Carnegie Mellon University Pittsburgh, PA, 1995.

[18] K.-K. Sung, T. Poggio, Example-based learning for view-based human face detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1) (1998) 39–51.

[19] R. Benenson, M. Omran, J. Hosang, , B. Schiele, Ten years of pedestrian detection, what have we learned?, in: ECCV, CVRSUAD workshop, 2014.

[20] C. Q. Lai, S. S. Teoh, A review on pedestrian detection techniques based on histogram of oriented gradient feature, in: Research and Development (SCOReD), IEEE, 2014, pp. 1–6.

[21] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.

[22] N. Inoue, T. Saito, K. Shinoda, S. Furui, High-level feature extraction using sift gmms and audio models, in: 20th International Conference on Pattern Recognition, IEEE, 2010, pp. 3220–3223.

[23] Q. Zhu, M.-C. Yeh, K.-T. Cheng, S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2006, pp. 1491–1498.

[24] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (12) (2006) 2037–2041.

[25] H. Jin, Q. Liu, H. Lu, X. Tong, Face detection using improved lbp under bayesian framework, in: IEEE First Symposium on Multi-Agent Security and Survivability, 2004, pp. 306–309.

[26] S. Liao, A. C. Chung, Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude, in: 8th Asian Conference on Computer Vision, 2007, pp. 672–679.

[27] T. Abegaz, G. V. Dozier, K. S. Bryant, J. Adams, B. Baker, J. Shelton, K. Ricanek, D. L. Woodard, Genetic-based selection and weighting for lbp, olbp, and eigenface feature extraction., in: 2011 Midwest Artificial Intelligence and Cognitive Science Conference, Special Session on Artificial Intelligence in Biometrics and Identity Sciences, CEUR, 2011, pp. 221–224.

[28] Z. Guo, L. Zhang, D. Zhang, Rotation invariant texture classification using lbp variance (lbpv) with global matching, Pattern Recognition 43 (3) (2010) 706–719.

[29] S. Munder, D. M. Gavrila, An experimental study on pedestrian classification, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (11) (2006) 1863–1868.

[30] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognitionf, 1997, pp. 130–136.

[31] C. Papageorgiou, T. Poggio, A trainable system for object detection, International Journal of Computer Vision 38 (1) (2000) 15–33.

[32] P. Felzenszwalb, D. McAllester, D. Ramanan, A discriminatively trained, multiscale, deformable part model, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[33] P. F. Felzenszwalb, D. McAllester, Object detection grammars, in: International Conference on Computer Vision, Workshops, 2011, p. 691.

[34] R. B. Girshick, P. F. Felzenszwalb, D. A. Mcallester, Object detection with grammar models, in: Neural Information Processing Systems, 2011, pp. 442–450.

[35] R. B. Girshick, P. F. Adviser-Felzenszwalb, From rigid templates to grammars: Object detection with structured models, in: Ph.D. dissertation, University of Chicago, 2012.

[36] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Visual object detection with deformable part models, Communications of the ACM 56 (9) (2013) 97–105.

[37] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, Cascade object detection with deformable part models, in: IEEE conference on Computer Vision and Pattern Recognition, 2010, pp. 2241–2248.

[38] S. Paisitkriangkrai, C. Shen, A. van den Hengel, Strengthening the effectiveness of pedestrian detection with spatially pooled features, in: European Conference on Computer Vision, Springer, 2014, pp. 546–561.

[39] W. Ouyang, X. Wang, Joint deep learning for pedestrian detection, in: International Conference on Computer Vision, 2013.

[40] X. Zeng, W. Ouyang, M. Wang, X. Wang, Deep learning of scene-specific classifier for pedestrian detection, in: European Conference on Computer Vision, Springer, 2014, pp. 472–487.

[41] S. Maji, A. C. Berg, J. Malik, Efficient classification for additive kernel svms, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 66–77.

[42] J. Wu, Efficient hik svm learning for image classification, IEEE Transactions on Image Processing 21 (10) (2012) 4442–4453.

[43] P. Viola, M. J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.

[44] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.