# Pedestrian Detection and Tracking using HOG and Oriented-LBP Features

Yingdong Ma, Xiankai Chen, and George Chen

Center for Digital Media Computing,
Shenzhen Institutes of Advanced Technology, Shenzhen, China

**Abstract.** During the last decade, various successful human detection methods have been developed. However, most of these methods are focused on finding powerful features or classifiers to obtain high detection rate. In this work we introduce a pedestrian detection and tracking system to extract and track human objectives using an on board monocular camera. The system is composed of three stages. A pedestrian detector, which is based on the non-overlap HOG feature and an Oriented LBP feature, is applied to find possible locations of humans. Then an object validation step verifies detection results and rejects false positives by using a temporal coherence condition. Finally, Kalman filtering is used to track detected pedestrians. For a $320 \times 240$ image, the implementation of the proposed system runs at about 14 frames/second, while maintaining an human detection rate similar to existing methods.

**Keywords:** Pedestrian detection, support vector machine, Oriented Local Binary Pattern, Histograms of Oriented Gradient

## 1 Introduction

Pedestrian detection has attracted considerable attention from the computer vision community over the past few years. One of the important reasons is its wide variety of applications, such as video surveillance, robotics, and intelligent transportation systems. However, detecting humans in video streams is a difficult task because of the various appearances caused by different clothing, pose and illumination. Moving cameras and cluttered background make the problem even harder.

Many human detection methods have been developed but most of these methods are focus on finding powerful features or classifiers to obtain high detection rate. For applications such as on-line human detection for robotics and automotive safety, both efficiency and accuracy are important issues that should be considered carefully. In this work, we study the issue of finding a feature set for human detection from onboard video streams. In particular it combines the non-overlap histograms of oriented gradient (HOG) appearance descriptor [1] and an oriented Local Binary Patterns (LBP) feature. Temporal coherence condition is employed to reject false positives from detection results and Kalman filtering is used to track detected pedestrians. The aim of the proposed system is to achieve accurate human detection, while maintains efficient for applications that require fast human detection and tracking.

The paper is structured as follows. Section II briefly reviews some recent works in human detection in static and moving images. Section III describes the proposed features and Section IV provides a description of the object validation and object tracking system. The implementation and comparative results are presented in Section V. Finally, Section VI summarizes the results.

## 2 Previous Work

Within the last decade a number of pedestrian detection systems have been presented to tackle the problem of finding humans from a moving platform. In this section we briefly review some more recent works on pedestrian detection. Systematic overviews of related work in this area can be found in [2] and [3].

While some pedestrian detection approaches are based on key-point detectors [4] or use a parts-based approach [5, 6], most up-to-date human detection approaches make use of the sliding-window analysis scheme. The performance of a sliding-window based method can be influenced by choosing various features and classifiers. Some widely used features extracted from the raw image data include Haar wavelet [7], HOG [1], edge orientation histogram (EOH) [8], edgelet [9], shapelet [10], region covariance [11], and LBP [12]. The most common classifiers, those employ statistical learning techniques to map from features to the likelihood of a pedestrian being present, usually either some variant of boosting algorithms [13, 14] or some types of support vector machines [15]. Researchers also explore different ways to combine various features to improve detection accuracy, such as the combination of gradient feature, edgelet, and Haar wavelets in [16] and the combined feature pool (Haar wavelets, EOH, edge density) in [13].

As an important visual feature, motion descriptors are widely used in video-based person detectors. Viola et al. [17] build a detector for static camera video surveillance by applying extended Haar wavelets over two consecutive frames of a video sequence to obtain motion and appearance information. In order to use motion for human detection from moving cameras, motion features derived from optic flow such as histograms of flow (HOF) are proposed by Dalal et al. [18]. These motion features are widely used in recent pedestrian detection systems [19, 20].

## 3 Feature Pool and Classifier

The proposed system aims to extract and track human objectives from onboard video streams. Efficiency and accuracy are two important issues of the system performance. In the following we describe the employed features including our proposed oriented Local Binary Patterns (Oriented LBP) feature and the HOG feature. This section also describes the classifier which we deployed in the sliding-window based system.

### 3.1 Feature Pool

Feature extraction is the first step in most object detection and tracking applications. The performance of these applications often relies on the extracted features. As mentioned above, a wide range of features has been proposed for pedestrian detection. We tried different successful features and their combination to choose suitable features for our moving camera pedestrian detection system. In particular we evaluate HOG, HOF, region covariance, LBP, and the color co-occurrence histograms [19]. The HOG and a new LBP features are chose because of the following reasons.

Firstly, the motion information is not included in the proposed system because the global motion caused by moving camera cannot be eliminated efficiently. The changing background also generates a large optical flow variance. Secondly, calculation of the region covariance feature and the color co-occurrence histograms (CH) is a time consuming task. For example, CH tracks the number of pairs of certain color pixels that occur at a specific distance. For a fixed distance interval and a quantized $n_c$ representative colors, there are $n_c(n_c+1)/2$ possible unique, non-ordered color pairs with corresponding bins in the CH. That is, in the case of $n_c$=128, CH has 8128 dimensions [19].

**HOG** Histograms of oriented gradients, proposed by Dalal and Triggs [1], are one of the most successful features in pedestrian detection applications. HOG features encode high frequency gradient information. Each $64\times128$ detection window is divided into $8\times8$ pixel cells and each group of $2\times2$ cells constitute a block with a stride step of 8 pixels in both horizontal and vertical directions. Each cell consists of a 9-bin histogram of oriented gradients, whereas each block contains a 36-D concatenated vector of all its cells and normalized to an $L^2$ unit length. A detection window is represented by $7\times15$ blocks, giving a total of 3780-D feature vector per detection window.

Although dense HOG features achieve good results in pedestrian detection, processing a $320\times240$ scale-space image still requires about 140ms on a personal computer with 3.0GHz CPU and 2GB memory. Hence, in our experiments we compute histograms with 9 bins on cells of $8\times8$ pixels. Block size is $2\times2$ cells with non-overlap (stride step 16 pixels). Each $64\times128$ detection window is represented by $4\times8$ blocks, yielding a total of 1152-D feature vector per detection window. According to [1], large stride step might decrease system performance. However, in our experiments, combining with other complementary features can significantly improve the system performance (see Fig. 2).

**Oriented LBP** As a discriminative local descriptor, LBP is originally introduced in [22] and shows great success in human detection applications [12]. LBP feature has several advantages such as it can filter out noisy background using the concept of uniform pattern [12] and it is computational efficiency. To calculate the LBP feature, the detection window is divided into blocks and computes a histogram over each block according to the intensity difference between a center pixel and its neighbors. The histograms of the LBP patterns from all blocks are then concatenated to describe the

texture of the detection window. For a 64×128 detection window with 32 non-overlap 16×16 blocks, its LBP feature has a 1888-D feature vector.

The HOG feature can be seen as an oriented gradient based human shape descriptor, while LBP feature serves as a local texture descriptor. Recent researches have shown that combination of these two features can achieve very good results in pedestrian detection [21]. However, extraction of the HOG-LBP feature is computational expensive. Each 64×128 detection window has a 5668-D (3780+1888) feature vector .

In this work we introduce a lower-dimensional variant of LBP, namely the oriented LBP. We define the arch of a pixel as all continuous "1" bits of its neighbors. The orientation $\theta(x,y)$ and magnitude $m(x,y)$ of a pixel is defined as its arch principle direction and the number of "1" bits in its arch, respectively (see Fig.1). The pixel orientation is evenly divided into k bins over $0^o$ to $360^o$. Then, the orientation histograms $F_{i,k}$ in each orientation bin k of cell $C_i$ are obtained by summing all the pixel magnitudes whose orientations belong to bin k in $C_i$

$$F_{i,k} = \sum_{\substack{(x,y)\in C_i \\ \theta(x,y)\in bin_k}} m(x,y) \qquad (1)$$

In our implementation, k is 8 for $LBP_{8,1}$, $C_i$ has the size of 8×8 pixels. In this way, a 64×128 detection window with 32 non-overlap 16×16 blocks has a 1024-D (4×8×32) oriented LBP feature vector. Finally, we have a 2176-D (1152+1024) HOG-Oriented LBP feature vector for each detection window. Fig. 1 illustrates the computation of Oriented LBP feature.

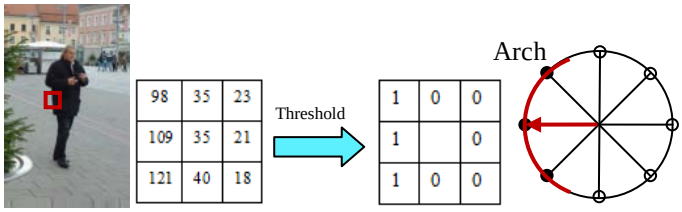



**Fig. 1**. Computing pixel orientation and magnitude of Oriented LBP feature. In this example we use a threshold of 20.

### 3.2 Classifier

Most pedestrian detection systems choose either SVMs or Adaboost as classifiers. We evaluate these classifiers using various common features and find the linear SVMs has better detection rate than that of boosting algorithms (see Fig.2 (b)). Moreover, with the lower-dimensional feature vector of the HOG-Oriented LBP feature, the processing speed of the same video frame using linear SVMs is similar to Adaboost. Therefore, we choose linear SVMs as classifier in the proposed pedestrian detection system.

# 4 Pedestrian Tracking

The output of the pedestrian detection step is a set of independent bounding boxes show possible locations of human objectives. Due to the cluttered background and limited number of positive/negative samples, the detection might have some false alarms. In order to recover from these problems, we employ a detection validation step before pedestrian tracking.

## 4.1 Detection Validation

Small objects with their height less than 40 pixels are discarded. After that, we compute a confidence measurement for each detected object based on its distance to the hyperplane of the SVM classifier. The distance between an example and the hyperplane can be calculated as follow:

$$d(x_i) = \frac{sgn(\omega x+b)}{|\omega|} = sgn(\sum_{j=1}^{i} y_j \alpha_j K(x_j, x_i + b)/|\omega| \quad (2)$$

where $K(x_j, x_i)$ is the kernel function.

The confidence measure of a pedestrian detection is in direct proportion to its distance to the hyperplane. Therefore, the confidence measure of an example is computed as:

$$conf(x_i) = \rho exp\ (-1/|d(x_i)|) \quad (3)$$

where $\rho$ is a normalizing factor. Objects with their confidence measure less than $\tau$ are discarded as false positives. In practice, we found that setting the threshold as $\tau \in [0.65, 0.75]$ can provide good results.

In the next step, bounding boxes that do not satisfy the temporal coherence condition are removed. We define this condition as follow. When the first object is detected, a Kalman filter is initialized to start pretracking. The Kalman filter predicts its location in the next frame. A new detection in a consequent frame is assigned to this track if it coherently overlaps with the tracker prediction. In practice we set the overlap rate as 0.7. Only candidates meeting this condition in three consecutive frames are considered as a stable pedestrian objective and are labeled as positive.

## 4.2 Pedestrian tracking

Once a candidate is validated as a pedestrian, pretracking stops and pedestrian tracking starts. In the object tracking step, each newly detected pedestrian with a positive mark is tracked by an individual Kalman filter. The detection validation and pedestrian tracking steps efficiently remove false positives from detection results.

## 5 Experimental Results

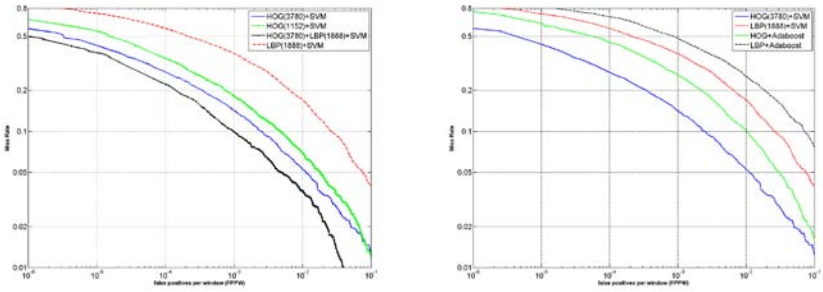
The proposed system is implemented on a personal computer with 3.0GHz CPU and 2GB memory running the Windows XP operating system and OpenCV libraries. For 320×240 pixel images, the implementation of the proposed system runs at about 13 to 15 frames/second, depending on the number of pedestrians being tracked.

We created several training and test video sequences containing thousands of positive (pedestrians) and negative (non-pedestrians) samples in different situations. The well-known INRIA person dataset is employed to evaluate the performance of the HOG-Oriented LBP feature. We also compare the detection results between the proposed system and several common pedestrian detection methods on both the INRIA dataset and the created video streams.

### 5.1 Performance of Different Features with SVM Classifier

First we implement the Dalal and Triggs algorithm using the same dataset and the same parameters suggested in their paper [1]. We compare its performance with other features, including non-overlap HOG, LBP, and HOG-LBP, using a linear SVMs classifier. As shown in Fig.2 (a), the HOG-LBP outperforms other features. However, the best performance of HOG-LBP is obtained by using a more complicated feature. As a result, the processing time of a 320×240 pixel video stream is about 7.7 frames per second (fps) using the HOG-LBP feature, whereas the cell-structured LBP feature has the fastest processing speed. The processing speed of various features on the same video stream using a linear SVMs classifier is shown in Table 1.

In our experiment, the HOG feature has block spacing stride of 8 pixels and the value of non-overlap HOG feature is 16 pixels. As Fig.2 (a) shows, overlapping blocks introduce redundant in the final descriptor vector but increase the performance. The miss rate decreases by 3% at $10^{-3}$FPPW when we change from HOG feature to non-overlap HOG feature. The advantage of non-overlap HOG is its high processing speed, about 40% faster than that of HOG feature.




(a) (b)

Fig.2. Performance comparison of different features (a) and different classifiers (b)

**Table 1.** Precessing speed of different features

| *Features* | *HOG* | *Non-overlap HOG* | *LBP* | *HOG-LBP* | *Non-overlap HOG-Oriented LBP* |
|---|---|---|---|---|---|
| ***Precessing time (fps)*** | **11.8** | **16.0** | **17.4** | **7.7** | **14.1** |

### 5.2 Comparison of SVM and Adaboost Classifier

SVM and Adaboost are the two most popular classifiers and are widely used in various pedestrian detection systems. We evaluate the performance of these classifiers using Haar wavelet, HOG, and LBP features on the INRIA dataset.

As shown in Fig.2 (b), using HOG and LBP features with SVM classifier outperforms Adaboost classifier on the INRIA dataset. We observe that the performance of Haar wavelet feature is worse than HOG and LBP features, which reflects the fact that the intensity pattern of human face is simple than that of human body. Hence, Haar wavelet feature is more suitable for human face detection applications.

By treating each bin of HOG feature as an individual feature, we implement HOG-Adaboost pedestrian detection on the INRIA dataset in order to comparing its performance to HOG-SVM detector. The block size changes from $12\times12$ to $64\times128$. We observe performance decrease by about 9% at $10^{-3}$FPPW when we change from SVM detector to Adaboost detector. Moreover, processing time of the two classifiers is similar when using the HOG feature. This is caused by the reason that computing histograms of oriented gradient for a sub-window spends much more time than computing intensity difference, even with the help of cascade structure and the integral images.

### 5.3 Detection results with HOG-Oriented LBP Feature

As our main contribution is integrating the non-overlap HOG feature with the Oriented LBP feature to achieve efficient and accurate human detection, we compare the performance of our non-overlap HOG-Oriented LBP feature with the HOG-LBP feature on the INRIA dataset. As we can see from Fig.3, the HOG-LBP feature achieves detection rate about 91% at $10^{-3}$FPPW, better than the proposed non-overlap HOG-Oriented LBP feature by about 3%. However, the processing speed of our proposed feature is about 2 times faster than the HOG-LBP feature.

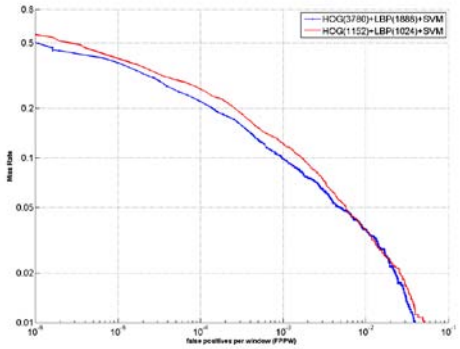



Fig.3. Performance comparison between HOG-LBP and the proposed feature

Fig.4 shows some pedestrian detection results of the augmented system on the INRIA dataset and our video sequences. Some examples with false positives are shown in the bottom row. As mentioned in section IV, using the detection validation method can efficiently remove most of these false alarms.

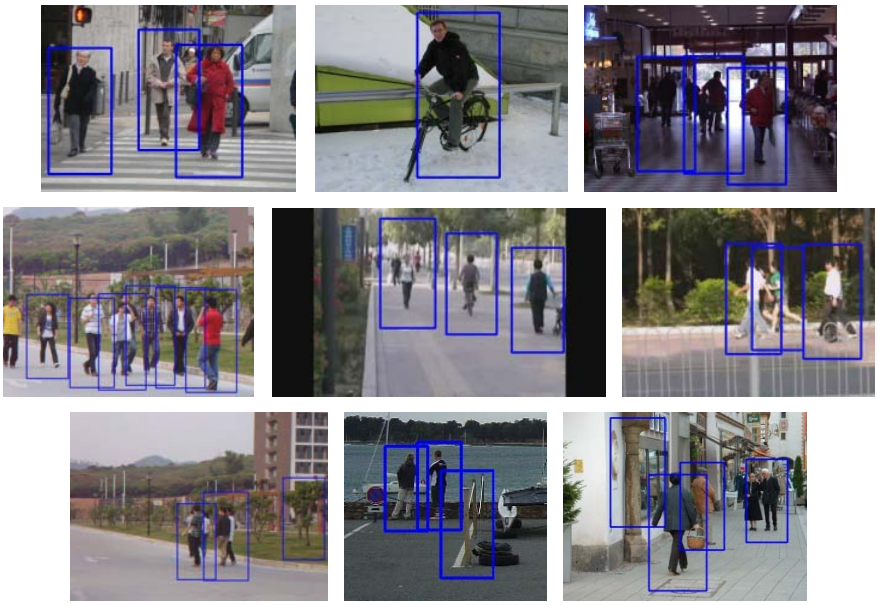


Fig.4. Pedestrian detection examples using the proposed system; Top row: examples from INRIA dataset; Middle row: examples of video streams; Bottom row: examples with false alarms

## 6 Conclusions

We introduce a new pedestrian detection system in this work, which aims at extracting and tracking human objectives from video streams with high efficiency and accuracy. We demonstrate the proposed human detection algorithm that has similar detection rate of up-to-date methods with an up to 2 times speedup. This is achieved by integrating the non-overlap HOG feature with an Oriented LBP feature. In this way, a lower dimensional and high discriminative feature vector is obtained for each detection window. Detection validation based on temporal coherence condition is employed to reject possible false alarms.

**Acknowledgments.** This work was supported in part by the NSFC research project (Grant No. 61003297).

## References

1. N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, vol.1, pp. 886-893, 2005
2. D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, “Survey of Pedestrian Detection for Advanced Driver Assistance Systems,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, No.7, pp.1239-1258, 2010
3. M. Enzweiler, and D. M. Gavrila, “Monocular Pedestrian Detection: Survey and Experiments,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, No.12, pp.2179-2195, 2009
4. E. Seemann, M. Fritz, and B. Schiele, “Towards robust pedestrian detection in crowded image sequences,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2007
5. I. P. Alonso, D. F. Llorca, M. A. Sotelo, L. M. Bergasa, P. R. Toro, M. Ocana, and M. A. G. Garrido, “Combination of Feature Extraction Methods for SVM Pedestrian Detection,” IEEE Transactions on Intelligent Transportation Systems, vol.8, No.2, pp. 292-307, 2007
6. B. Wu and R. Vevatia, “Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors,” IEEE International Conference on Computer Vision (ICCV), Vol. 1, pp. 90-97, 2005
7. C. Papageorgiou and T. Poggio, “A trainable system for object detection,” International Journal of Computer Vision, 38(1): 15-33, 2000
8. K. Levi and Y. Weiss, “Learning Object Detection from a Small Number of Examples: the Importance of Good Features,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 53-60, 2004
9. B. Wu, and R. Nevatia, “Detection and Segmentation of Multiple, Partially Occluded Objects by Grouping, Merging, Assigning Part Detection Responses,” International Journal of Compute Vision vol. 82, pp. 185-204, 2009
10. P. Sabzmeydani and G. Mori, “Detecting Pedestrians by Learning Shapelet Features,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2007
11. S. Paisitkriangkrai, C. Shen, and J. Zhang, “Fast Pedestrian Detection Using a Cascade of Boosted Covariance Features,” IEEE Transactions on Circuits and Systems for Video Technology, vol.18, No.8, pp. 1140-1151, 2008

12. Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou, “Discriminative Local Binary Patterns for Human Detection in Personal Album,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1-8, 2008
13. Y. Chen and C. Chen, “Fast Human Detection Using a Novel Boosted Cascading Structure With Meta Stages,” IEEE Transactions on Image Processing, vol.17, No.8, pp. 1452-1464, 2008
14. T. Kim and R. Cipolla, “MCBoost: Multiple Classifier Boosting for Perceptual Co-clustering of Images and Visual Features,” Neural Information Processing Systems Foundation, NIPS, 2008
15. Z. Lin and L. S. Davis, “A pose-invariant descriptor for human detection and segmentation,” European Conference on Computer Vision, ECCV, 2008
16. B. Hu, S. Wang, and X. Ding, “Multi Features Combination for Pedestrian Detection,” Journal of Multimedia, vol. 5, No. 1, pp. 79-84, 2010
17. P. Viola, M. J. Jones, and D. Snow, “Detecting Pedestrians Using Patterns of Motion and Appearance,” IEEE International Conference on Computer Vision, ICCV, 2003
18. N. Dalal, B. Triggs, and C. Schmid, “Human Detection Using Oriented Histograms of Flow and Appearance,” ECCV, Part II, LNCS 3952, pp. 428-441, 2006
19. S. Walk, N. Majer, K. Schindler, and B. Schiele, “New Features and Insights for Pedestrian Detection,” IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1030-1037, 2010
20. C. Wojek, S. Walk, and B. Schiele, “Multi-Cue Onboard Pedestrian Detection,” IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 794-801, 2009
21. X. Wang, T. X. Han, and S. Yan, “A HOG-LBP Human Detector with Partial Occlusion Handling,” IEEE International Conference on Computer Vision, ICCV, 2009
22. T. Ojala, M. Pietik¨ainen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1):51–59, 1996