# Pedestrian Detection Based on Multi-Block Local Binary Pattern and Biologically Inspired Feature

Aminou Halidou[1], Xinge You[1] & Bachirou Bogno[2]

[1] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China

[2] Department of Physics, Higher Teachers' Training College, University of Maroua, Cameroon

Correspondence: Aminou Halidou, Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Hongshan, Wuhan, Hubei 430074, China. Tel: 86-134-1969-5312. E-mail: aminouhalidou@gmail.com

## Abstract

Nowadays pedestrian detection plays an important role in security and driving assistance. Detecting moving object is complex, and some of the detection methods are comparatively ineffective and slow. In relation to human detection it is very useful to combine independent information sources, such as appearance and motion. To achieve acceptable detection performance, we propose inter-frames differencing image to compute the region of interest, and MB-BIF to extract features. The MB-BIF approach combines two well-known methods, the Multi-Block Local Binary Pattern and Biologically Inspired Method. We evaluate the performance of different features descriptors on different databases, and our method shows good efficiency.

**Keywords:** differencing image, biologically inspired feature, Multi-Block Local Binary Pattern analysis, enhanced Fisher Linear Discriminant

## 1. Introduction

In recent years, there has been a great progress in detection systems. Lot of efforts have been made in the development of pedestrian detection systems, for its importance in practical applications such as, driving safety and surveillance systems. Detecting moving pedestrian is important in video surveillance. In general the motion detection is based on three approaches (Radke, Andra, Al-Kofahi, & Roysam, 2005), known as optical flow (Velastin, Boghossian, Lo, Sun, & Vicencio-Silva, 2005) based on the intensity change of the moving objects, background subtraction (Piccardi, 2004) based on the difference between the given frame and the background model, frame differencing (Kim & Hwang, 2002) based on simple consecutive frames difference. These approaches treat pedestrian detection as two categories: human and non-human classification problems. There is a need for an automatic system that will detect a person in a complex scene; this step is crucial for human movements. Although a significant number of approaches have been developed, the detection of pedestrian is far from being put in place (Sharma & Davis, 2007). The complexity of human pedestrian detection is due to the wide variation in complexion, backgrounds and environmental conditions. Therefore, a big concern is given to the implementation of a system capable of differentiating humans from a captured image with high accuracy, precision, and efficiency.

Pedestrian detection has been extensively studied, significant work has been devoted on detecting, locating, and tracking people in images and videos. Stereovision was the first approach used for human detection in the late 1990s based on Fundamental Matrix algorithm (Franke & Kutzbach, 1996). In 1998, Heisele and Woehler developed an algorithm based on the movement of legs compared to the ground to detect and classify individual (Heisele & Woehler, 1998). Haar-like features and support vector machine (SVM) is proposed in the same year (Papageorgiou, Oren, & Poggio, 1998). In year 2000, researchers started taking advantage of the face detections algorithms. In 2001, due to the face detection success, the human detection work speeded up, many works are based on face detection technics such the Viola and Jones method which gives a good result (Viola & Jones, 2001b). Three year later, LBP (Local binary patterns) is introduced in the same area of face detection (Ahonen, Hadid, & Pietikainen, 2004). In 2005, HOG (Histogram of Oriented Gradient) algorithm is introduced by the researchers of

INRIA (Institut National de Recherche en Informatique et en Automatique), it has been proven to be efficient (Dalal & Triggs, 2005). In 2006, credit was given to the covariance descriptor (Tuzel, Porikli, & Meer, 2006). In 2007, the shape features is introduced to describe local pieces (head, shoulder, and body) (Sabzmeydani & Mori, 2007). In 2008 researchers at Rutgers University improved the work by using the covariance matrix as descriptor (Tuzel, Porikli, & Meer, 2008). A significant performance is obtained by combining different pedestrian feature extraction approaches and classification algorithms (Paisitkriangkrai, Shen, & Zhang, 2008).

The key step in image recognition and video analysis is the feature extraction from the region of interest (ROI). This paper proposes a pedestrian detector inspired from human visual system with a hierarchical architecture topographical. The hierarchical architecture topographical maps the input image to the feature-extraction layers. Where, the last layer performs the classification as pedestrian or not. In order to solve the problems encountered in standard moving object detection methods we proposed a Fast determination of ROI and MB-BIF features.

The paper is organized as follows: Section 2 mentions the related work. Section 3 presents the framework of our approach. Section 4 describes the detection procedure. Section 5 analyzes the experimental results and Section 6 draws conclusion.

## 2. Related Work

Pedestrian detection has gained an extensive amount of interest in computer vision community over past years. Here we present some motion detector approaches, Broggi, Fascioli, Carletti, Graf and Meinecke (2004) used multi-resolution texture symmetry, edge symmetry and edge density ROI extraction. In Liu and Fujimura (2004), strategies are to apply intensity threshold, and the motion constraint. In Xu, Liu, and Fujimura (2005), intensity thresholds are used with the combination of support-vector-machine classifier and Kalman-Filter. The sliding window approaches are more promising when the image resolution is low. The first sliding window was proposed by Papa Georgiou, it was applied with the support vector machine to an over-complete dictionary of multi-scale Haar wavelets (Papageorgiou & Poggio, 2000). The histogram of oriented gradient (HOG) introduced by Dalal and Triggs (Dalal & Triggs, 2005) gained popularity over intensity based features. The Haar-like features proposed by Viola and Jones (Viola & Jones, 2001a), achieved a real-time performance by introducing integral image and multistage classifiers structure. The combination of many features increases the detector efficiency, (Wu & Nevatia, 2008) combined HOG, edgelet and covariance features. The multilayer network is widely used in pattern recognition, such as face detection (Garcia & Delakis, 2004), handwritten digit recognition (LeCun, Bottou, Bengio, & Haffner, 1998), facial expression analysis (Fasel, 2002). Hubel and Wiesel (1965) introduced Neural Network ideas, where local features are extracted following certain hierarchical structure. The network architectures (Fukushima, 2007) is made of many stages that contain different layers in turn, called the S-layer (simple layer), the C-layer (complex layer) and the V-layer. Dollar, Wojek, Schiele and Perona (2012) present a full understanding of the pedestrian detection. This work introduced a new region of interest determination, efficient features extraction, and powerful classification.

## 3. Our Approach

Figure 1 presents our system overview, the input video and images are preprocessed using video denoising (Zuo, Liu, Tan, Wang, & Zhang, 2013) and image denoising (Gupta, Mahle, & Shriwas, 2013) respectively before being convert into image frames. Section 3.1 describes how the ROI is computed from consecutive frames, section 3.2 presents our novel features descriptor based on Multi-block Bio-Inspired Features (MB-BIF), and section 3.3 indicates the classification process.

### 3.1 ROI Determination

In normal pedestrian detection, the detector searches all possible sub-windows of an image to find whether they contain a pedestrian or not, but this technique is computationally expensive. Our ROI extraction will therefore narrow down the searching region. Differencing Image technique detects consistent and reliable region of interest with bright intensity contrasts. Figure 2 presents the flowchart of the region of interest computation process.

Comparing two adjacent frames we can detect the dissimilarities and extract features from their differences. Taking two frames at time $T$, and $T + t$, using point A $(x, y)$ we can compute the differencing image as:

$$D_{Fimg(T,T+t)}(x,y) = \begin{cases} 0 & if\ (|Fimg_T(x,y) - Fimg_{T+t}(x,y)|) < Th \\ 1 & if\ (|Fimg_T(x,y) - Fimg_{T+t}(x,y)|) \geq Th \end{cases} \tag{1}$$
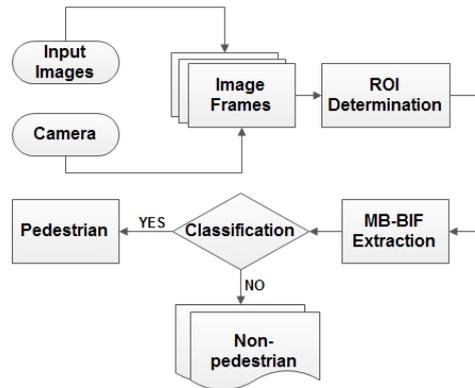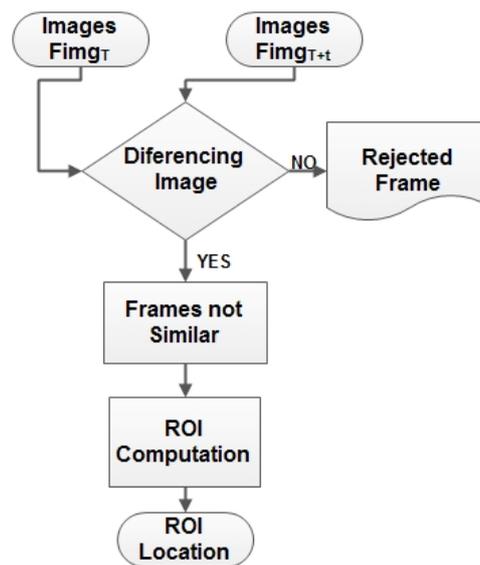
Figure 1. System overview



Figure 2. ROIs computation

Where $Fimg_T$ is the current video frame image, $t$ is the time interval between two consecutive frames, $Th$ the predefined threshold. In the proposed Differencing Image algorithm below $T_{end}$ indicates the stopping time, and $n$ defines step value to control the time interval assigned manually. The step (2) controls the ending process while the step (3) detects the frames similarities. If the similarity is beyond the threshold the concern frame can be skipped and next one is processed.

---

**Algorithm 1** Inter frames differencing image algorithm

---

    Initial parameters t = 1 and $n = 0, ..., T_{end}$
2:  **for** $T < T_{end}$  **do**
     Compute the Differencing Image $D_{Fimg(T,T+t)}$
4:     **if** $(D_{Fimg(T,T+t)} = 1)$ **then**
       Extract regions coordinates
6:     **end if**
     set t = $t + n$
8:  **end for**

---

From two consecutives frames of Figure 3 (a-b), the ROI is extracted based on the method given above. The Figure 3 (c) indicates the position of moving image elements near the camera, moving objects beyond 20 meters are not considered by the system. The differencing image is affected by the crowdedness of moving objects and the frames

interval times. The longer interval time (few frames per second) produce a distinctive region of interest. Figure 3 (d) shows the computed ROI in which our MB-BIF descriptor will extract individuals from other moving objects.

| (a) Fame 1 | (b) Fame 2 | (c) Moving Object | (d) ROI Location |

Figure 3. Region of moving objects computation

*3.2 Multi-Block Bio-Inspired Features (MB-BIF)*

In this sub-section, we introduce the MB-BIF descriptor for pedestrian detection. The MB-BIF framework contains four layers. The L1 layer is built of filters, the L2 layer extract nonlinear features, L3 layer uses MB-BIF to extract features, L4 layer reduces the dimension and improves the discrimination power of features.

3.2.1 L1 Layer

Gabor filter focus on local areas to extract information of precise scale and orientation, with certain orientation and scale it yields edges and bars along the direction, and extract facts in the corresponding frequency band. Therefore, Gabor filter extracts added facts in some significant body areas, which are very useful for pedestrian representation. Here we analyze the input image by an array corresponding to the primary visual cortex ($V$) cells. Due to its good cortical model (Guo, Mu, Fu, & Huang, 2009), the Gabor filter is applied in L1. Considering the image $I(x, y)$ with width $w$ and height $h$, the convolutional equation is given as:

$$G_I \psi(x, y, \mu, \nu) = I(x, y) * \psi_{\mu,\nu}(z) \tag{2}$$

With $\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|}{\sigma^2} e(-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{\sigma^2})[e^{ik_{u,\nu}z} - e^{\frac{-\sigma^2}{2}}]$ and $k_{\mu,\nu} = k_\nu e^{i\phi_\mu}$, $k_\nu = 2^{-\frac{\nu=2}{2}}\pi$, $\phi_u = u\frac{\pi}{8}$, $u = 0, 1..., U-1$, $v = 0, 1..., V-1$, Where $\mu$ and $\nu$ are the scales and orientations parameters, respectively. There are 8 bands (16 scales) in normal bio-inspired features, 2 filters for each band with variable number of orientations (4, 8, 16). Most detection and recognition systems applied 128 Gabor Magnitude Pictures (GMPs) with 8 bands and 8 orientations, here we use the same strategy. The GMPs is: $\{G_{ij}^q; i = 0, ..., m; j = 0, ..., n; q = 0, 1\}$ where $G_{ij}^0$ And $G_{ij}^1$ are filtered values at band i and orientation j, m is number of bands, n is orientations number, q is the scale index. Taken an image size of $64 \times 128$, we get 8192 dimensions from each GMP, and the output dimensions of L1 layer is $(64 \times 128 \times 16 \times 8) = 1048576$.

3.2.2 L2 Layer

From the previous layer output, a maximum scale is filtered out from two consecutive scales and computed as:

$$L2_{ij} = \max(G_{ij}^0, G_{ij}^1) \tag{3}$$

$L2_{ij}$ the maximum value of two consecutive L1 units at band $i$ and orientation $j$. the output of the L2 layer in turn is computed as:

$$L2 = (L2_{11}, L2_{12}, ..., L2_{1n}, L2_{21}, L2_{22}, ..., L2_{mn}) \tag{4}$$

3.2.3 L3 Layer

Here, we introduce the Multi-block Local Binary Pattern Bio-inspired Features (MB-LBP). MB-LBP features are extracted instead of common Haar-like features, rapidly calculated through the integral image. The extracted features are few with rich information and distinctive performance. Here the simple difference method in Haar-like features is replaced by encoding rectangular region by local binary pattern operator. The original LBP operator considers measurements from a 3x3 pixel square (Ojala, Pietikainen, & Harwood, 1996). The binary code is computed for every pixel by thresholding the 3 x 3 neighborhood pixel value with the center value and taking the result as a binary number. The multi-block LBP follows the rule of LBP operator, by encoding rectangle areas,
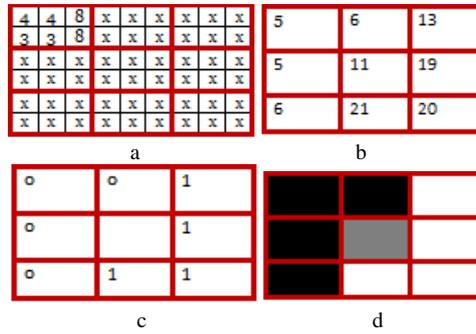
Figure 4. MB-LBP

Figure 4.a represents a region of an image divided into rectangles, and then the average value of each rectangle is calculated in Figure 4.b. The MB-LBP operator is defined by comparing central rectangle value with its neighborhood rectangles to provide a binary sequence.

$$\text{MB-LBP operator}: MB - LBP = \sum_{i=1}^{8} S\left(g_i - g_c\right) 2^i \tag{5}$$

Where $g_c$ is the central rectangle and $g_i$ (i = 0 .....8) are neighborhood rectangles.

$$S(x) = \begin{cases} 0, & \text{if } x > 0 \\ 1, & \text{if } x \leq 0 \end{cases} \tag{6}$$

As the original Local Binary Pattern, MB-LBP can capture a bunch of structures, 256 kinds of binary patterns are computed but not all of them are used to represent body shape, see Figure 5.
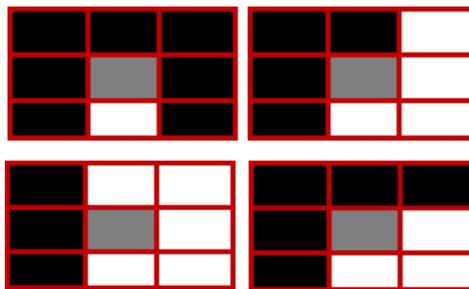
Figure 5. Samples of MB-LBP features

While extracting the MB-LBP features, a number of overlapping blocks are counted repeatedly. To increase the extraction efficiency and avoid counting repetition problem, the integral image is computed. The integral image provides a fastest way of computing the first two steps of Figure 4. The integral image is an algorithm that quickly and efficiently computes the sum of values in a rectangular subset of a grid (Crow, 1984). As the name is saying, the value at any point (x, y) in the summed area table is just the sum of all the pixels above and to the left of (x, y), only.

$$I(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \tag{7}$$

The computation of the summed area table is performed in a single pass over the image, using the fact that the value in the summed area table at (x, y) is just:

$$i(x, y) = [I(x, y) + I(x - 1, y - 1)] - [I(x - 1, y) + I(x, y - 1)] \tag{8}$$

After the summation, evaluating any rectangle can be accomplished in constant time with just four array references as seen in Figure 6.

$$\sum_{\substack{A(x)<x'\leq C(x)\\A(y)<y'\leq C(y)}} i(x',y') = [I(N) + I(K)] - [I(L) + I(M)] \tag{9}$$
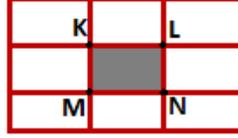


Figure 6. Finding the sum of a rectangular area

3.2.4 L4 Layer

The feature vector of each pedestrian image is formed using the concatenated L3 features. L4 layer applied Principal Component Analysis (PCA) to reduce the dimension of the resultant features and enhance the discriminative ability. In PCA, the projection matrix $W$ is composed by the orthogonal eigenvectors of the covariance matrix of all the training samples. To further improve the dimension reduction, we turn to the feature vector partitioning. The output L of the L3 layer can be rewritten as:

$$L3 = (H_1, H_2, ..., H_k) \tag{10}$$

where $H_n$ is the $n^{th}$ feature segment with a specific number of spatial histograms. For each feature segment $H_n$ a PCA model is computed for low-dimensional representation in the subspace:

$$L_d = W_d^T H_d \tag{11}$$

Thus, in MB-BIF, by building $k$ subspaces, the input pedestrian image is finally represented as:

$$L_d MB - BIF = (L_{d1}, L_{d2}, ..., L_{dk}) \tag{12}$$

*3.3 Enhanced Fisher Linear Discriminant Analysis*

Introduced by Fisher (1936) ensures the optimal separation in the new subspace by searching the optimal projection vector $W_{opt}$ from n samples in the mapped 1-D domain.

$$J(\Phi) = \Phi^T S_B \Phi / \Phi^T S_W \Phi \tag{13}$$

$S_B = \sum_{i=1}^{c} n_i(u_i - \overline{x})(u_i - \overline{x})^T$ is the between classes scatter matrix. The between classes function characterizes the discriminant structure of data, it makes the data points of different classes as distant as possible. $S_W = \sum_{i=1}^{c} \sum_{x_k \epsilon class_i} (u_i - \overline{x})(u_i - \overline{x})^T$ is the within classes scatter matrix. The within classes function characterizes the similarity between data from the same class, it makes the data points from the same class as close as possible. Where, $\Phi$ is $n$ dimensional vector, $u_i$ is the mean of the $i$ class, $c$ the number of the classes, $n_i$ is the quantity of samples of the $i$ class and $x_k$ the $k-th$ sample, $\overline{x}$ is the total sample mean of the vector. When $W_{opt}$ is maximum $\Phi$ is identically to the optimal projection vector $[V, D] = eig S_W^{-1} S_B$. Where $V$ and $D$ represent the eigenvalue and the eigenvector of $S_w^{-1} S_B$ respectively and $W_{opt}$ computed as $D$ correspond to maximum $V$. The fisher classifier is:

$$F(X) = \begin{cases} class1, & (W_{opt})^T > \theta \\ class2, & (W_{opt})^T \leq \theta \end{cases} \tag{14}$$

$W_{opt}$ is the optimal projection vector, $x$ is the input sample, $\theta = \frac{n_1 \widetilde{u}_1 + n_2 \widetilde{u}_2}{n_1 + n_2}$ is the threshold, $n_1$ is the number of samples with class $i$, $\widetilde{u}_1$ is the mean of samples with classes $i$. The within classes function is prone to map some points, that are not very close to each other, to a subspace in which they are very close to each other. This indicates that the variation of the values of data from the same class is impaired and lost in the reduced space space (Weinberger, Packer, & Saul, 2005). The enhanced discriminant analysis incorporates the variation of data from

the same class into the Fisher criterion to avoid the over-fitting problem in the reduced space (Gao, Liu, Zhang, Hou, & Yang, 2012). The discriminant feature extraction criterion can be rewritten as:

$$W_{opt} = \arg\max_{w} \frac{\Phi^T[\varepsilon S_B + (1 - \varepsilon)S_D]\Phi}{\Phi^T S_W \Phi} \tag{15}$$

Where $S_D$ stands for variability scatter matrix, $\varepsilon$ is a constant used to control the balance between discriminative ability and stableness of the algorithm.

## 4. Pedestrian Detection

This section describes the proposed approach for pedestrian detection. It starts with the ROIs, follows by MB-BIF, and ends with EFLDA. A pedestrian is considered as detected if the candidate image pass the entire test.

### 4.1 Dataset

We evaluate our work on the TUD-Brussels dataset, Caltech pedestrian benchmark. The TUD-Brussels contains pedestrians from various scales and viewpoints. It comes with 508 images of 640 x 480 pixels, 1326 annotated pedestrians, and 192 negative images. The Caltech Pedestrians is largely publically available pedestrian dataset. It offers a large number of samples. The annotated dataset is near to 2.5 hours of video captured at 30fps. It contents: 11 segmented sessions, 6 for training (S0 to S5) and 5 for testing (S6 to S10). The individual in these datasets appear in many positions, orientations, and background variety.

### 4.2 Region of Interest

The regions of interests are computed to speed up the detection system. The strategy consists to filter out unwanted image regions. Using differencing image histogram, pedestrians positions are computed on significant motion (presented by closest picks in histogram).

### 4.3 The Features Extraction

For MB-LBP feature, the image model of (Kuo, Yang, & Yen, 2012) serves to construct an effective auxiliary in the determination of a MB-BIF feature extraction frame. Figure 7 shows a pedestrian featured image of $64 \times 128$ pixels segmented into $8 \times 16$ square cells in $8 \times 8$ pixels. The model is built as the average pixel values in each square cell. The MB-LBP filters are used to recognize the distinctive features and their position.
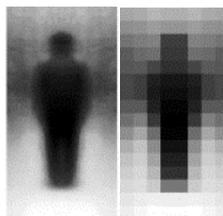


Figure 7. Pedestrian model

### 4.4 Training Algorithm

The detection accuracy depends on the training strategies, although the good combination of features and classifiers can bring progress in performance, that progress is much less obvious than increasing the number of training examples. Therefore, a correct training process is another possibility to improve the detection performance.

---

**Algorithm 2** The MB-LBP training algorithm

---
1: Integral image computation from negative and positive input images
2: Pedestrian region selection within the positive samples
3: Arbitrary region selection within the negative samples
4: Evaluate the effective features from the large feature set
5: Steps 1 to 4 is repeated till the needed numbers of training samples are collected
6: EFLDA is performed for classification

---

*4.5 Detection Process*

In the preprocessing step the region of interest is located. The detection window is applied to the candidate region given by the preprocessing step, features are extracted and then submitted to the classifier for test, will be considered as pedestrian the result that passes all the steps. The information concerning that window is recorded to mark the region of detected pedestrian.

## 5. Experiment and Results

This part presents the experimental results and analysis. We ran the experiment on 2.2 GHz Intel Core 2 Duo T6600 with 4 GB of memory using Mat lab Release R2011b. Figure 8 presents features a performances result, MB-BIF approach shows a good performance, followed by region covariance descriptor. The performances are evaluated using detection rate versus false positive rate. Figure 9 presents the detected pedestrians in red color bounding box.
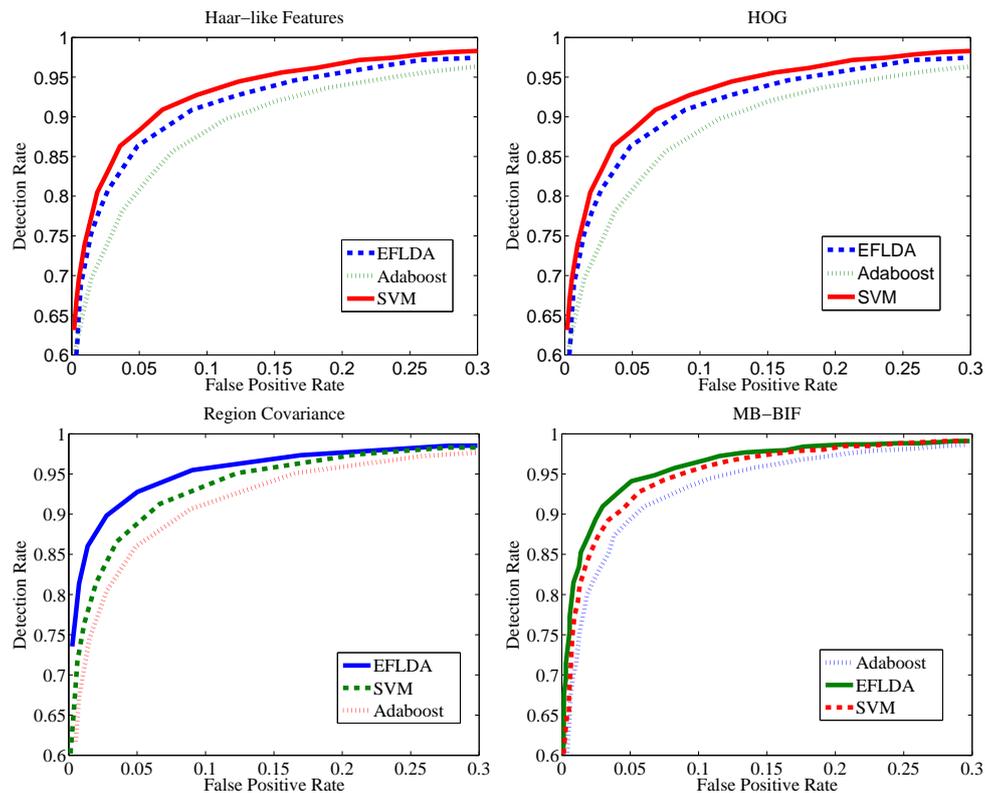


Figure 8. Performance evaluation



Figure 9. Detected pedestrians

Table 1 presents the MB-BIF average result using TUD-Brussels Pedestrian and Caltech Pedestrian dataset. The

detection is counted as correct, if it overlaps the annotated by 70% at least, using intersection-over-union measurement $PC(r, o) = |r \bigcap o|/|r \bigcup o|$ .

Table 1. Accuracy of our detection system

| Datasets | Samples | Detected | Ratio |
|---|---|---|---|
| TUD-Brussels Pedestrian | 1326 Pos | 1291 | 97.36% |
| TUD-Brussels Pedestrian | 0192 Neg | 14 | 7.03% |
| Caltech Pedestrian | Video 1 | 63 | 91.3% |
| Caltech Pedestrian | Video 2 | 47 | 94% |
| Caltech Pedestrian | Video 3 | 33 | 91.5% |
| Caltech Pedestrian | Video 4 | 58 | 95% |

Table 2 presents the processing speed of our approach on our recorded video, the image size of 720x480 has a speed of 15fps, the image size of 640x480 has a speed of 22 fps, and the image size of 320x240 has a speed of 71fps. The whole system worked with an average speed of 33 fps. We also noticed that the detection time increases with the size of the image as well as with the number of pedestrians in the image.

Table 2. The processing speed of our system

| Image Size | Our Method |
|---|---|
| 720x480 | ~ 15 fps |
| 640x480 | ~ 22 fps |
| 320x240 | ~ 71 fps |

## 6. Conclusion

Motivated by the success of MB-LBP and BIF, we used them for the moving pedestrian detection. Firstly, the ROI determination is filtering out similar frames and unwanted sub-windows to reduce the computational time. Secondly, the combination of the MB-LBP and the BIF extract features from candidate region. Thirdly, a high discriminative classification is obtained using EFLDA preceded by PCA features reduction. The experimental results demonstrated the high performance and effectiveness of our approach. Our approach is fast enough to analyze high video frame rate using less memory. Extensive experimental results illustrate the advantages of the proposed method.

Although the person detectors techniques have proven a reliable and good detection, there is still place for real time detection improvements. In the future work, we wish to extend the current approach for handicapped pedestrian detection, and include some novel classifiers and feature descriptors to optimize the detector.

## References

Ahonen, T., Hadid, A., & Pietikainen, M. (2004). Face recognition with local binary patterns. In *Computer vision-eccv 2004* (pp. 469-481). Springer. http://dx.doi.org/10.1007/978-3-540-24670-1_36

Broggi, A., Fascioli, A., Carletti, M., Graf, T., & Meinecke, M. (2004). A multi-resolution approach for infrared vision-based pedestrian detection. In *Intelligent vehicles symposium, 2004 ieee* (pp. 7-12).

Crow, F. (1984). Summed-area tables for texture mapping. *Computer Graphics, 18*(3), 207-212. http://dx.doi.org/10.1145/964965.808600

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer vision and pattern recognition, 2005. IEEE computer society conference on* (Vol. 1, pp. 886-893).

Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: An evaluation of the state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34*(4), 743-761. http://dx.doi.org/10.1109/TPAMI.2011.155

Fasel, B. (2002). Robust face analysis using convolutional neural networks. In *Pattern recognition, 2002. proceedings. 16th international conference on* (Vol. 2, pp. 40-43).

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics, 7*(2), 179-188.

Franke, U., & Kutzbach, I. (1996). Fast stereo based object detection for stop amp;go traffic. In *Intelligent vehicles symposium, 1996., proceedings of the 1996 IEEE* (p. 339-344).

Fukushima, K. (2007). Interpolating vectors for robust pattern recognition. *Neural Networks, 20*(8), 904-916. http://dx.doi.org/10.1016/j.neunet.2007.06.003

Gao, Q., Liu, J., Zhang, H., Hou, J., & Yang, X. (2012). Enhanced fisher discriminant criterion for image recognition. *Pattern Recognition*. http://dx.doi.org/10.1016/j.patcog.2012.03.024

Garcia, C., & Delakis, M. (2004). Convolutional face finder: A neural architecture for fast and robust face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26*(11), 1408-1423. http://dx.doi.org/10.1109/TPAMI.2004.97

Guo, G., Mu, G., Fu, Y., & Huang, T. S. (2009). Human age estimation using bio-inspired features. In *Computer vision and pattern recognition, 2009. IEEE conference on* (pp. 112-119).

Gupta, V., Mahle, R., & Shriwas, R. S. (2013). Image denoising using wavelet transform method. In *Wireless and optical communications networks (wocn), 2013 tenth international conference on* (pp. 1-4). http://dx.doi.org/10.1109/WOCN.2013.6616235

Heisele, B., & Woehler, C. (1998). Motion-based recognition of pedestrians. In *Pattern recognition, 1998. proceedings. fourteenth international conference on* (Vol. 2, pp. 1325-1330).

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology*.

Kim, C., & Hwang, J. N. (2002). Fast and automatic video object segmentation and tracking for content-based applications. *Circuits and Systems for Video Technology, IEEE Transactions on, 12*(2), 122-129. http://dx.doi.org/10.1109/76.988659

Kuo, Y., Yang, Z., & Yen, C. (2012). Fast pedestrian detection system with a two layer cascade of classifiers. *Computers & Mathematics with Applications, 64*(5), 1311-1323. http://dx.doi.org/10.1016/j.camwa.2012.03.075

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324. http://dx.doi.org/10.1109/5.726791

Liu, X., & Fujimura, K. (2004). Pedestrian detection using stereo night vision. *Vehicular Technology, IEEE Transactions on, 53*(6), 1657-1665. http://dx.doi.org/10.1109/TVT.2004.834876

Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition, 29*(1), 51-59. http://dx.doi.org/10.1016/0031-3203(95)00067-4

Paisitkriangkrai, S., Shen, C., & Zhang, J. (2008). Fast pedestrian detection using a cascade of boosted covariance features. *Circuits and Systems for Video Technology, IEEE Transactions on, 18*(8), 1140-1151. http://dx.doi.org/10.1109/TCSVT.2008.928213

Papageorgiou, C., Oren, M., & Poggio, T. (1998). A general framework for object detection. In *Computer vision, 1998. sixth international conference on* (pp. 555-562).

Papageorgiou,C., & Poggio, T. (2000). A trainable system for object detection. *International Journal of Computer Vision, 38*(1), 15-33. http://dx.doi.org/10.1023/A:1008162616689

Piccardi, M. (2004). Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on* (Vol. 4, pp. 3099-3104).

Radke, R. J., Andra, S., Al-Kofahi, O., & Roysam, B. (2005). Image change detection algorithms: a systematic survey. *Image Processing, IEEE Transactions on, 14*(3), 294-307. http://dx.doi.org/10.1109/TIP.2004.838698

Sabzmeydani, P., & Mori, G. (2007). Detecting pedestrians by learning shapelet features. In *Computer vision and pattern recognition, 2007. IEEE conference on* (pp. 1-8). http://dx.doi.org/10.1109/CVPR.2007.383134

Sharma, V., & Davis, J. (2007). Integrating appearance and motion cues for simultaneous detection and segmentation of pedestrians. In *Computer vision, 2007. IEEE 11th international conference on* (pp. 1-8).

Tuzel, O., Porikli, F., & Meer, P. (2006). Region covariance: A fast descriptor for detection and classification. *Computer Vision-ECCV 2006* (pp. 589-600). http://dx.doi.org/10.1007/11744047_45

Tuzel, O., Porikli, F., & Meer, P. (2008). Pedestrian detection via classification on riemannian manifolds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30*(10), 1713-1727. http://dx.doi.org/10.1109/TPAMI.2008.75

Velastin, S. A., Boghossian, B. A., Lo, B. L., Sun, J., & Vicencio-Silva,M. A. (2005). Prismatica: toward ambient intelligence in public transport environments. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 35*(1), 164-182. http://dx.doi.org/10.1109/TSMCA.2004.838461

Viola, P., & Jones, M. (2001a). Rapid object detection using a boosted cascade of simple features. In *Computer vision and pattern recognition, 2001. Proceedings of the 2001 IEEE computer society conference on* (Vol. 1, pp. I-511).

Viola, P., & Jones, M. (2001b). Robust real-time object detection. *International Journal of Computer Vision, 4.*

Weinberger, K., Packer, B., & Saul, L. (2005). Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization. In *Proceedings of the tenth international workshop on artificial intelligence and statistics* (pp. 381-388).

Wu, B., & Nevatia, R. (2008). Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection. In *Computer vision and pattern recognition, 2008. cvpr 2008. IEEE conference on* (pp. 1-8).

Xu, F., Liu, X., & Fujimura, K. (2005). Pedestrian detection and tracking with night vision. *Intelligent Transportation Systems, IEEE Transactions on, 6*(1), 63-71. http://dx.doi.org/10.1109/TITS.2004.838222

Zuo, C., Liu, Y., Tan, X., Wang, W., & Zhang, M. (2013). Video denoising based on a spatiotemporal kalmanbilateral mixture model. *The Scientific World Journal*, 2013. http://dx.doi.org/10.1155/2013/438147

## Copyrights