

PEDESTRIAN DETECTION BASED ON FOREGROUND SEGMENTATION IN VIDEO SURVEILLANCE

¹MING XIN, ²CAILI FANG

¹Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475004, China

²School of Computer and Information Engineering, Henan University, Kaifeng 475004, China

E-mail: ¹xinming@henu.edu.cn, ²fcl@henu.edu.cn

ABSTRACT

Real-time detection and tracking of moving pedestrians in image sequences is a fundamental task in many computation vision applications such as automated visual surveillance system. In this paper we propose a human detection method based on foreground segmentation, and the detection speed is satisfying for the application of video surveillance. During detection, unlike the exhaustive scan typically used in general human detection systems, in order to avoid scanning regions like the sky, the foreground segmentation stage is firstly implemented in the video surveillance sequences by utilizing Gaussian mixture model algorithm, and then, human detection stage is executed on the regions of interest (ROI) extracted from the video surveillance sequence frame. In contrast with the exhaustive scan without explicit segmentation, our proposed approach can meet the real-time requirement. Carefully designed experiments demonstrate the superiority of our proposed approach.

Keywords: *Pedestrian Detection, Object Tracking, Foreground Segmentation*

1. INTRODUCTION

Effective techniques for pedestrian detection are of special interest in computer vision since many applications involve people's locating and tracking. Thus, significant research has been devoted to detecting, locating and tracking people in images and videos. Both pedestrian tracking and pedestrian locating are the problem of finding the positions of all pedestrian objects in a video surveillance sequence. More specifically, the goal is to find the bounding box for each pedestrian object in video surveillance application.

Considerable approaches to pedestrian detection have been explored over the last few years. One common approach is to use a sliding window to scan the image exhaustively in scale-space, and classify each candidate window individually [1,2], for which only a binary classifier is needed to indicate whether a target is pedestrian or not from a given window. However, this procedure has two main shortcomings: 1) many irrelevant regions are passed to the next module (e.g., sky regions or ROIs inconsistent with perspective), which increases the potential number of false positives. 2) The number of candidates is large, which makes it difficult to meet real-time requirements, although

some proposals have recently studied this problem [3,4].

To meet the real-time requirement in video surveillance application, variable improved approaches are explored in recent years. Zhang et al. [5] propose a multi-resolution framework to reduce the computational cost. Leibe et al. [6] present a technique termed as the implicit shape model. The idea is to use a keypoints detector, Hessian- Laplace in this case, then compute a shape context descriptor for each keypoint, and finally, cluster them to construct a codebook. During recognition, each detected keypoint is matched to a cluster, which then votes for an object hypothesis using Hough voting, thus avoiding a candidate generation step. Zhu et al. [7] propose a rejection cascade using HOG features. Also based on HOG features, Begard et al. [8] address the problem of real-time pedestrian detection by considering different implementations of the AdaBoost algorithm.

In general, the pose of surveillance camera does not change. According to the properties, the foreground segmentation stage is firstly implemented in the video surveillance sequences by utilizing Gaussian mixture model algorithm. Therefore, the candidate pedestrian regions are

firstly obtained. There are twofold benefits with respect to the following pedestrian detection stage: 1) The reduction of the image area to be processed means that the overall system time can also be reduced; 2) by submitting fewer background regions for classification, the rate of false alarms can also be reduced while the same detection rate is maintained.

Foreground segmentation and pedestrian detection do not exist in isolation. Both of them can

provide useful prior or posterior information to each other. The foreground segmentation stage can effectively reduce the detection area such as sky regions or ROIs inconsistent with perspective. Simultaneously, the pedestrian detection stage can predict future pedestrians' positions, and feed the foreground segmentation algorithm with pre-candidates. The algorithm flowchart is illustrated as Fig.1.

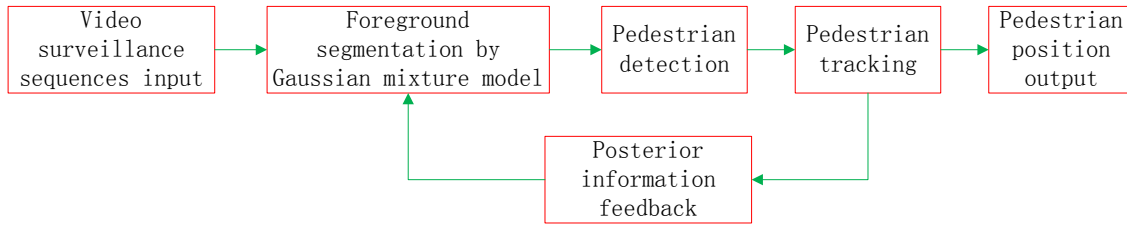


Fig.1 The Proposed Algorithm Flowchart

The remainder of this work is organized as follows. In Sec. 2, we review Gaussian mixture model algorithm briefly. Sec. 3.1 and Sec. 3.2 introduce the Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) features, respectively. Experimental results are given in Sec. 4, followed by conclusions in Sec. 5.

2. GAUSSIAN MIXTURE MODEL

Gaussian mixture model plays an important role in foreground segmentation fields due to its analytical tractability, asymptotic properties, and ease of implementation. For foreground segmentation, the basic Gaussian mixture model algorithm models each background pixel by a mixture of K Gaussian distributions. Different Gaussians are assumed to represent different colors. The weight parameters of the mixture represent the time proportions that those colors stay in the scene. The brief introduction is as follows:

Each pixel in the detection frame is modeled by a mixture of K Gaussian distributions. The probability that a certain pixel has a value of x_t at time t can be written as

$$p(x_t) = \sum_{j=1}^k w_j p(x_t; \theta_j) \tag{1}$$

where w_j is the weight subject to $w_j > 0$ and $\sum_{j=1}^k w_j = 1$. Each component's density

$p(x; \theta_j)$ is a normal probability distribution represented by

$$p(x; \theta_k) = p(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)} \tag{2}$$

where $\theta_k = \{\mu_k, \Sigma_k\}$ represents the k 'th component center and covariance, respectively. Given a group of N independent and identically distribution samples $X = \{x^1, x^2 \dots x^N\}$, the log-likelihood corresponding to a k -component mixture is

$$\log p(X; \theta) = \log \prod_{i=1}^N p(x^i; \theta) = \sum_{i=1}^N \log \sum_{j=1}^k w_j p(x^i; \theta_j) \tag{3}$$

Because the maximum likelihood estimate of θ cannot be solved analytically, Expectation Maximization algorithm has been widely used to estimate the parameters of Gaussian mixture model. It is an iterative algorithm, and the details were introduced in literatures [9];

3. HOG AND LBP FEATURES

Considerable literatures have shown that significant improvement in human detection can be achieved by combining different low-level features. A strong set of features provides high discriminatory power. In this section, we will introduce Histogram of Oriented Gradients (HOG) features and Local Binary Pattern (LBP) features, respectively.

3.1 Histogram Of Oriented Gradients

In literature [1,10], each detection window is represented as a set of overlapping blocks. Each block contains four non-overlapping regions called cells, figure 2 displays the detailed position relationship between cells and blocks.

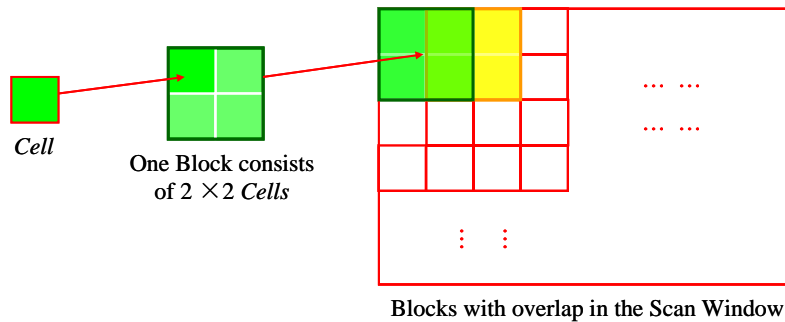


Fig. 2 The Position Relationship Between Cells And Blocks

The calculation process of HOG descriptor is as follows:

1. First, for each pixel of the cell, horizontal gradient G_x and vertical gradient G_y is calculation through gradient operator $[-1, 0, 1]$.

$$G_x(x, y) = I(x+1, y) - I(x-1, y)$$

$$G_y(x, y) = I(x, y+1) - I(x, y-1)$$

(4)

2. According the horizontal gradient G_x and vertical gradient G_y , for each pixel of the cell, the gradient orientation $O_g(x, y)$ and gradient values $V_g(x, y)$ are computed as

$$O_g(x, y) = \arctan(G_y(x, y) / G_x(x, y)) \quad (5)$$

$$V_g(x, y) = \sqrt{(G_x(x, y))^2 + (G_y(x, y))^2} \quad (6)$$

where, $O_g(x, y) \in (0^0, 180^0]$.

3. According to figure 2, obtaining the interesting region of cells and blocks.

4. Finally, each cell is represented as a 9-bin histogram with each bin corresponding to a particular gradient orientation, here, gradient orientation is divided into 9 bins, each bin includes 20^0 interval. Then, each block is represented by a 36-D unit-length vector. To reduce the influence of large gradient magnitudes, the L2-Hys normalization was applied. In this way, a 3780-D feature vector is used to represent a detection window.

3.2 Local Binary Pattern

The original LBP operator introduced in literature [11] labels the pixels of an image by thresholding the 3x3 neighborhood of each pixel with the center pixel value and considering the result as a binary number. Then the 256-bin histogram of the labels can be used as a texture descriptor. Fig.3 shows an example of LBP calculati

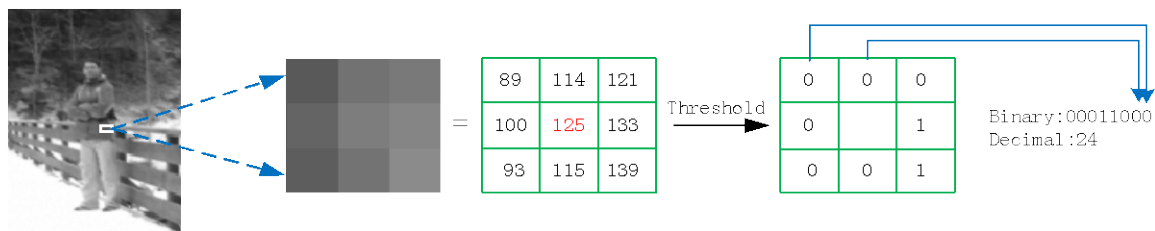


Fig. 3 The Basic LBP Operator

In order to deal with textures at different scales, the LBP operator is later extended to use neighborhoods of different sizes. For example, the operator $LBP_{4,1}$ uses 4 neighbors while $LBP_{16,2}$ considers the 16 neighbors on a circle of radius 2. In general, the operator $LBP_{P,R}$ refers to a neighborhood size of P equally spaced pixels on a circle of radius R that form a circularly symmetric neighbor set.

Another extension to the original operator is the definition of so called *uniform patterns*. Ojala et al. [11] defined these fundamental patterns as those with a small number of bitwise transitions from 0 to 1 and vice versa. For example, 00000000 and 11111111 contain 0 transition while 00000110 and 01111110 contain 2 transitions and so on. Accumulating the patterns which have more than 2 transitions into a single bin yields an LBP descriptor.

To consider the shape information of humans, we divide human images into M small non-overlapping

cells C_1, C_2, \dots, C_M . The LBP histograms extracted from each cell are then concatenated into a single, spatially enhanced feature histogram.

Since feature combination can significantly improve detector performance, we combine gradient based HOG feature and texture based LBP feature, and the proposed human detection algorithm achieves desiring performance. More detailed experimental results are shown in the next section.

4. EXPERIMENTAL RESULTS

To evaluate our proposed method, we have tested it on real video surveillance sequences. These sequences include objects such as pedestrian, cars etc. In general, pedestrians are interesting objects in video surveillance systems, and the main goal of surveillance system is to automatically detect and locate pedestrians in video sequences once they are initialized.

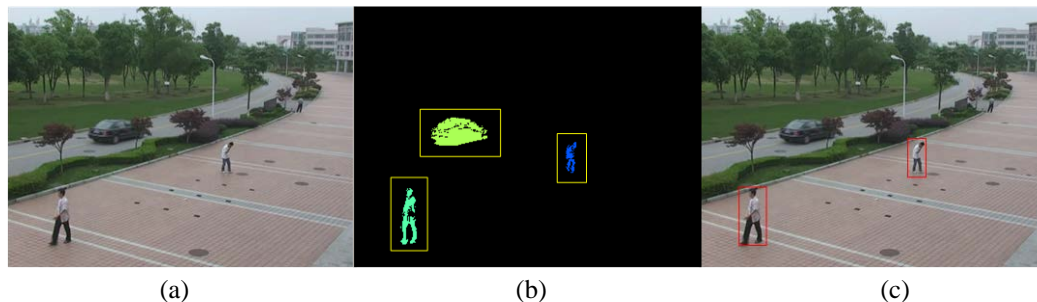


Fig.4 Pedestrian Detection Based On Foreground Segmentation

Fig.4a displays the input frame; Fig.4b presents the foreground segmentation results derived by Gaussian mixture model; Fig.4c demonstrates the pedestrian detection results by utilizing our proposed approach. Fig.4 illustrates that pedestrian detection algorithm this paper proposed can avoid scanning the total image by utilizing the foreground segmentation step, which submits fewer foreground regions for classification. The reduction of the image area to be processed means that the overall system time can also be reduced. According the detection results, our proposed algorithm can effectively detect the location and size of pedestrian by excluding the interference foreground region yielded by moving car.

5. CONCLUSIONS

This paper presents an effective pedestrian detection algorithm, and which can meet the real-time requirement in a certain extent. The general human detection systems adopt the exhaustive scan strategy, which makes it difficult to fulfill real-time requirements. In fact, just like our proposed approach, we can take advantage of some application prior knowledge so that the number of ROIs to process can be greatly reduced. However, there are still some drawbacks in this approach, such as the camera pose must be fixed. In the future work, we will try to explore more effective and real-time pedestrian detection algorithm.



ACKNOWLEDGMENTS

The work is supported by National Science Foundation of China (Grant No. 60972119);

REFERENCES:

- [1] N. Dalal and B. Triggs, Histogram of oriented gradient for human detection, In IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp.886-893.
- [2] P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision, 63(2), 2005, pp.153-161.
- [3] C. Wojek, G. Dorko´ , A. Schulz, and B. Schiele, Sliding-Windows for Rapid Object Class Localization: A Parallel Technique, Proc. Symp. German Assoc. for Pattern Recognition, 2008, pp.71-81.
- [4] W. Zhang, G. Zelinsky, and D. Samaras, Real-Time Accurate Object Detection Using Multiple Resolutions, Proc. Int’l Conf. Computer Vision, 2007, pp.1-8.
- [5] W. Zhang, G. Zelinsky, and D. Samaras. Real-time accurate object detection using multiple resolutions. In International Conference on Computer Vision, 2007.
- [6] B. Leibe, A. Leonardis, and B. Schiele, Robust Object Detection with Interleaved Categorization and Segmentation, International journal of computer vision, 77(1-3), 2008, pp.259-289.
- [7] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp.1491-1498.
- [8] J. Begard, N. Allezard, and P. Sayd. Real-time human detection in urban scenes: Local descriptors and classifiers selection with adaboost-like algorithms. In IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2008.
- [9] P. KadewTraKuPong and R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection. in Proc. 2nd European Workshp on Advanced Video-Based Surveillance Systems, 2001.
- [10] Y.-T. Chen and C.-S. Chen. Fast human detection using a novel boosted cascading structure with meta stages. IEEE Trans. on Image Processing, 17(8), 2008, pp.1452–1464.
- [11] Timo Ojala, Matti Pietikäinen, Topi Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 2002, pp.971–987.