

Robust Head-shoulder Detection by PCA-Based Multilevel HOG-LBP Detector for People Counting

Chengbin Zeng, Huadong Ma

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia

Beijing University of Posts and Telecommunications

Beijing, China

cbzeng@gmail.com, mhd@bupt.edu.cn

Abstract—Robustly counting the number of people for surveillance systems has widespread applications. In this paper, we propose a robust and rapid head-shoulder detector for people counting. By combining the multilevel HOG (Histograms of Oriented Gradients) with the multilevel LBP (Local Binary Pattern) as the feature set, we can detect the head-shoulders of people robustly, even though there are partial occlusions occurred. To further improve the detection performance, Principal Components Analysis (PCA) is used to reduce the dimension of the multilevel HOG-LBP feature set. Our experiments show that the PCA based multilevel HOG-LBP descriptors are more discriminative, more robust than the state-of-the-art algorithms. For the application of the real-time people-flow estimation, we also incorporate our detector into the particle filter tracking and achieve convincing accuracy.

Keyword: head-shoulder detection, multilevel HOG-LBP detector, PCA, people counting

I. INTRODUCTION

Counting the number of people for surveillance systems has widespread applications in public places, such as museums, markets, and scenic spots. For example, the public museum can control the amount of entering people according to the real-time people flow information. Therefore, it is necessary to develop an automatic system of estimating the number of people accurately and rapidly.

The computer vision based methods are widely used to people counting. Kilambi *et al* [1] proposed a blob-based system to estimate the number of people in a group in urban environments. However, the blob-based methods only can detect moving objects. Furthermore, when there are shadows of the people themselves, the count is affected greatly.

Li *et al* [2] applied the Histograms of Oriented Gradients (HOG) feature [3] to detect the head-shoulder. The merit of the head-shoulder detection is its effectiveness of reducing the partial occlusion. But the HOG feature is not discriminative enough for the head-shoulder detection [2, 4].

Wang *et al* [4] combined the HOG feature with the Local Binary Pattern (LBP) feature [5] to detect the holistic body of human. They also reported that the HOG-LBP feature gains more than 20% improvement for upper body detection compared to the HOG feature. But the detection rate still

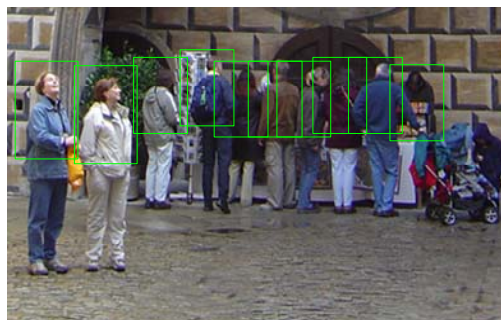


Figure 1. The performance of our head-shoulder detector in a complex scene. The image is scanned at 16 scales to search for head-shoulders of humans. We achieve minimal false alarms even though there are partial occlusion occurred (best visualized in green color).

cannot satisfy the demands of the people counting system (74% detection rate with a false alarm of 10^{-4}).

In this paper, we propose a robust and rapid head-shoulder detector for people counting. We first use the multilevel HOG-LBP feature to detect the head-shoulders of people, which can enhance the detection rate significantly. To further improve the detection accuracy, Principal Components Analysis (PCA) [6] is used to reduce the dimension of the multilevel HOG-LBP feature set.

Our experiments show that the PCA-based multilevel HOG-LBP descriptors are more discriminative, more robust than state-of-the-art algorithms. Figure 1 shows the detection result by using our method. For the application of the real-time people-flow estimation, we also incorporate our detector into the particle filter tracking [7] and increase the detection accuracy.

In the rest of the paper, we give a detailed description of our algorithm and experimental results.

II. OUR HEAD-SHOULDER DETECTOR

We use the multilevel HOG-LBP feature to detect the head-shoulder of the human. This is inspired by [8], which uses multi-scale histogram of oriented edge energy feature to detect the holistic body of pedestrians. As in [9] we use PCA to reduce the dimension of feature set. It is worth men-

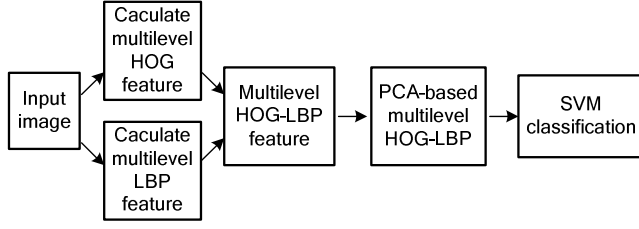


Figure 2. The process of our head-shoulder detection

tioning that Felzenszwalb *et al.* [10] used PCA-HOG to detect deformable objects. Figure 2 shows the process of our head-shoulder detection algorithms based on the PCA-HOG-LBP detector.

A. Multilevel HOG feature

HOG [3] is an excellent descriptor for capturing the edge direction or the distribution of local intensity gradients of objects. It has been applied successfully to detect the holistic body of the human.

In order to improve the performance of HOG, we use multilevel HOG feature to describe the head-shoulder. Figure 3 illustrates the procedure of the feature extraction. We first compute the gradient magnitude (Figure 3 (b)) of the input image using 1-D masks, *i.e.*, $[-1 \ 0 \ 1]$. The size of the input image is 48×64 pixels with gamma normalization.

Second, we divide the gradient magnitude of the image into blocks at three levels (Figure 3(c)). Each block at each level consists of four rectangle cells. The gradient magnitude of each pixel in the cell is voted into 9 bins according to the orientation of the pixel's gradient. The nine orientation bins are evenly spaced over 0° - 180° .

Third, the histograms of the four cells in each block (Figure 3(d)) are concatenated into a feature vector of the block. The L2-Hys normalization [3] of the feature vector is used to reduce the influence of the local variation in illumination and foreground-background contrast. Then, the feature vectors of the blocks are concatenated into the feature vector of each level.

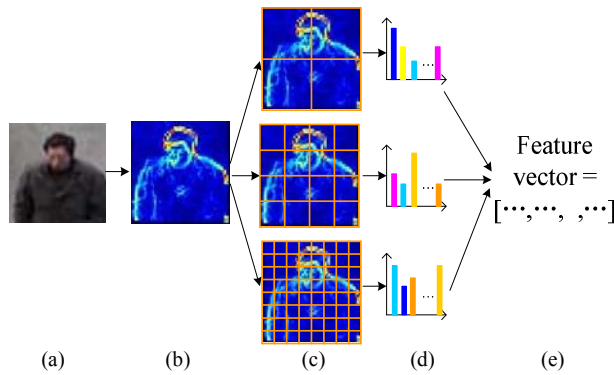


Figure 3. The procedure of the feature extraction of the multilevel HOG. (a) The input image. (b) The gradient magnitude of the input image. (c) The block division at three levels. (d) The histograms of each level. (e) The final multilevel HOG feature.

Finally, three feature vectors corresponding to three levels are concatenated into the final 3024-D multilevel HOG feature vector (Figure 3(e)).

B. Multilevel LBP Feature

As we mentioned in Section 1, HOG, as an edge descriptor, is not discriminative enough for head-shoulder detection. The combination of the edge shape information and the texture information will enhance the detection performance significantly. Thus, as in [4] we use LBP [5] as the texture descriptors of the head-shoulder.

The LBP is an excellent texture descriptor for its invariance to gray-scale and rotation. We extend the work of [11], and use multilevel block structured LBP to describe the head-shoulders of people. The same as Figure 3(c), we divide the gray image of the input image into blocks at three levels. Then, we calculate the histograms of the LBP patterns for each block. The LBP patterns we used is $LBP_{8,1}^2$ (Figure 4), which denotes that 8 points with radius 1 are sampled for each pixel, and the number of 0 - 1 transitions is at most 2. A binary pattern is called uniform pattern if the binary pattern contains at most two 0 - 1 transitions.

For each block at one level, pixels in the block with different uniform patterns are voted into different bins and all of the nonuniform patterns are voted into one bin. We then use L2-Hys normalization scheme for the histograms of the blocks, which is better performance than L1-sqrt normalization scheme used in [4] (by 5% with a false alarm of 10^{-4}). Finally, the three LBP feature vector corresponding to the three levels are concatenated into the final 4956-D multilevel LBP feature vector.

C. PCA-based Multilevel HOG-LBP Detector

By combining the multilevel HOG and the multilevel LBP as the feature set, we can detect the upper bodies of people robustly even though there partial occlusions occurred. To further improve the detection rate, we use PCA [6] to discover low dimensional feature of multilevel HOG-LBP.

PCA, as a standard technique for dimensionality reduci-

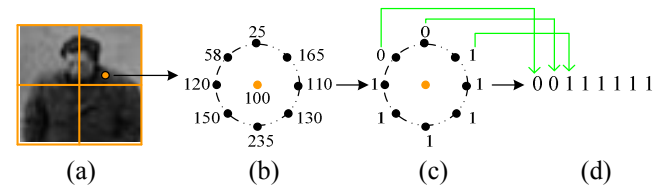


Figure 4. An illustration of the $LBP_{8,1}^2$ pattern. (a) The input gray image at one level. (b) 8 sampling points with radius 1 of the pixel in the block. Bilinear interpolation is used to obtain the pixel values when the sampling point is not in the center of a pixel. (c) The binary value of the sampling point taking the pixel value as the threshold. (d) The $LBP_{8,1}^2$ pattern of the pixel.

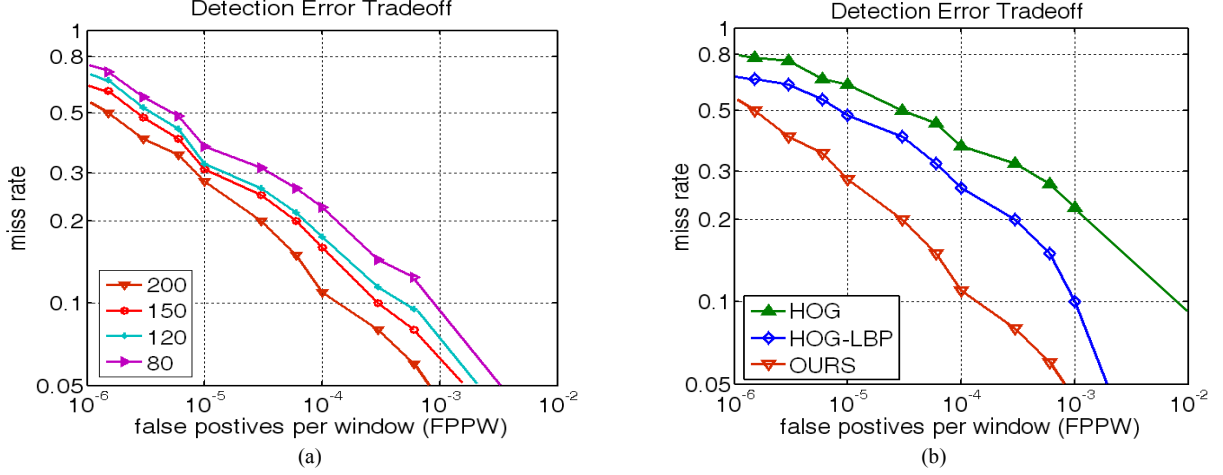


Figure 6. (a) The performance comparison of the multilevel HOG-LBP detectors with different dimensions projected by PCA. (b) The performance comparison between the proposed head-shoulder detectors and state-of-the-art detectors on our head-shoulder dataset.

on, has been widely applied in the area of computer vision such as face recognition. We use PCA to reveal the hidden structures of multilevel HOG-LBP feature by computing a linear transformation that maps the feature set from a high dimensional space to a lower dimensional space. This transformation can accurately represent the high dimensional feature with a compact feature representation. More importantly, projecting the multilevel HOG-LBP feature onto the low-dimensional space can discard the redundant and noisy information. Thus the PCA-based multilevel HOG-LBP detector makes the detection of the head-shoulder more robust.

III. EXPERIMENTS

A. Dataset

We created a head-shoulder dataset, which includes 2000 48×64 positive samples. Some of the positive samples are shown in Figure 5. About half of the positive samples are cropped from the INRIA dataset [3], and the others are cropped from the Internet. The 5000 negative examples are randomly sampled from person-free images. We also add some lower body images to the negative set to reduce false alarm.

B. Training

We use the linear SVM ($c=1$) to train and classify on the head-shoulder dataset. Hard examples (false positives) are necessarily added to the negative set to re-train and then obtain the final detector.



Figure 5. Some of the positive samples from our head-shoulder dataset.

C. Performance Evaluation

To quantify the detection performance, we use Detection Error Tradeoff (DET) curves [3], and plots of miss rate versus false positives per window (FPPW). Lower miss rate means better detection performance on the same FPPW.

We first evaluate the performance of the multilevel HOG-LBP detectors on different PCA subspaces. The PCA subspace with a dimensionality of 200 performs best (Figure 6(a)). Therefore, we use the PCA-based multilevel HOG-LBP with a dimensionality of 200 as our final head-shoulder detector. Then, we evaluate the HOG detector and HOG-LBP detector respectively on our head-shoulder dataset and achieve similar results as [4]. Finally, we test our PCA-based multilevel HOG-LBP detector and achieve 89% detection rate at 10^{-4} FPPW as shown in Figure 6(b). Compared to the HOG detector (62% detection rate at 10^{-4} FPPW) and the HOG-LBP detector (74% detection rate at 10^{-4} FPPW), our detector improves the detection rate significantly.

We present the detection results in Figure 7 for some surveillance images. The detection speed is less than 0.2 seconds for a 320×240 image (2.5 GHz dual core CPU and 2GB RAM).



Figure 7. Some of the head-shoulder detection results.

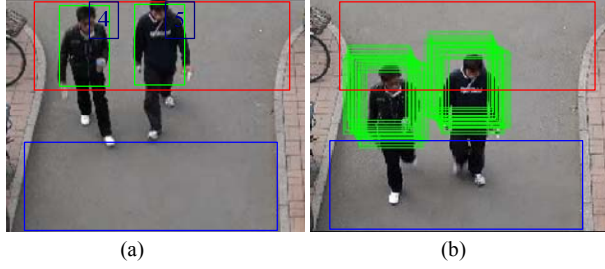


Figure 8. An illustration of our detection and tracking method. (a) Detect the head-shoulder of people in the region of red rectangle. (b) Tracking the detected targets. See text for details.

D. Counting the Pedestrian Flow

We apply our detector to the application of counting the pedestrian flow. We use the “detection and tracking” strategy by incorporating our detector into the particle filter tracking [7].

As illustrated in Figure 8, the red rectangle denotes the detection region. The interval of detection is 10 frames to avoid the repetitive detection. The detected targets are labeled a number (Figure 8(a)), and then be tracked by using particle filter with classifiers.

For the particle tracking, we use a 2-dimensional normal distribution to generate the initial particle set. As illustrated in Figure 8(b), each green rectangle represents a particle. The number of particles of each object is twenty. Then, we calculate the likelihood of each particle according to its PCA-based multilevel HOG-LBP feature. The rest steps of tracking are described in [7].

Once the targets being tracked enter the blue region (Figure 8), the amount of targets will be added to the counting system. Figure 9 illustrates the whole procedure of the people counting.

Our people counting system can run in real time (12 frames/second with 2.5 GHz dual core CPU and 2GB RAM) because the size of the detection region is only 64×180 pixels and the number of particles for each target are only twenty. Moreover, our system successfully detects and tracks 95% emerged targets.

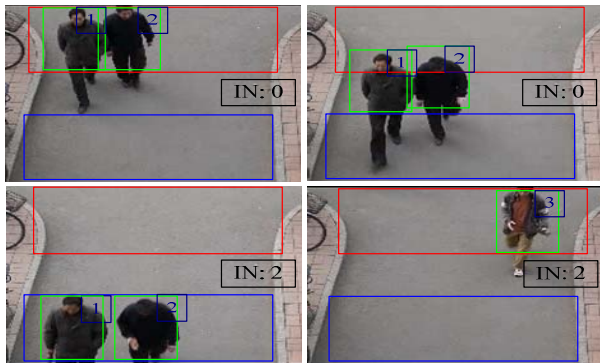


Figure 9. An illustration of our people counting system.

IV. CONCLUSION

We propose a novel head-shoulder detector which uses multilevel HOG-LBP as the feature set and uses PCA to discard the redundant and noisy information of the feature set. Compared to the state-of-the-art algorithms, our detector is more distinctive leading to significant improvements in detection accuracy. We also incorporate the detector into the particle filter tracking to count the pedestrian flow and achieve convincing accuracy.

ACKNOWLEDGMENT

This work reported in this paper is supported by the National Science Fund for Distinguished Young Scholars under Grant No.60925010, the National Natural Science Foundation of China under Grant No. 60833009, the 111 Project under Grant No.B08004, the National High Technology Research and Development Program of China under Grant No. 2009AA01Z305.

REFERENCES

- [1] P. Kilambi, E. Ribnick, and *etc.* Estimating pedestrian counts in groups. In *CVIU*, 2007.
- [2] M. Li, Z. X. Zhang, K. Q. Huang and T. N. Tan. Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection. In *ICPR*, 2008.
- [3] N.Dalal and B.Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.
- [4] X. Y. Wang, X. Han and S. C. Yan. An HOG-LBP human detector with partial occlusion handling. In *ICCV*, 2009.
- [5] T. Ojala, M. Pietikäinen and T. Mäenpää. Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. In *PAMI*, 2002.
- [6] I. T. Joliffe. Principal Component Analysis. Springer-Verlag, 1986
- [7] T. Chateau, V. Gay-Belille and *etc.* Real-Time tracking with Classifiers. In *ECCV*, 2006
- [8] S. Maji, A. C. Berg and J. Malik. Classification Using Intersection Kernel Support Vector Machines is efficient. In *CVPR*, 2008.
- [9] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *CVPR*, 2004.
- [10] P. F. Felzenszwalb, R. B. Grishick, D. McAllister, and D. Ramanan. Object detection with discriminatively trained part based models. In *PAMI*, 2009.
- [11] T. Ahonen, A. Hadid and M. Pietikäinen, Face description with local binary patterns: application to face recognition. In *PAMI*, 2006.