Automatic surveillance system using fish-eye lens camera

Xue Yuan (袁 雪)*, **Yongduan Song (宋永端)**, **and Xueye Wei (魏学业)**

Center for Intelligent Systems and Renewable Energy, Beijing Jiaotong University, Beijing 100044, China

**Corresponding author: xyuan@bjtu.edu.cn*

Received June 24, 2010; accepted October 21, 2010; posted online January 28, 2011

This letter presents an automatic surveillance system using fish-eye lens camera. Our system achieves wide-area automatic surveillance without a dead angle using only one camera. We propose a new human detection method to select the most adaptive classifier based on the locations of the human candidates. Human regions are detected from the fish-eye image effectively and are corrected for perspective versions. An experiment is performed on indoor video sequences with different illumination and crowded conditions, with results demonstrating the efficiency of our algorithm.

OCIS codes: 110.0110, 100.0100, 150.0150.

doi: 10.3788/COL201109.021101.

Due to large field of view, wide-angle lens are popularly used for various applications, such as surveillance, robotic navigation, and semi-automatic parking systems. Because the angle of view of the fish-eye lens used in our system was up to 185°, it achieved effective wide-area surveillance without a dead angle only one camera. However, it brought an inherent distortion in the image, and this distorted image must be rectified or restored in order to recognize and understand the image accurately. Human detection and tracking is a necessary approach for automatic surveillance systems. However, in the image taken by our surveillance system, the region where human enters the surveillance space is distorted and it is difficult to detect humans using the original method introduced in Refs. [1−9]. To our knowledge, there is still no reliable pedestrian detection algorithm reported for fish-eye image. Refernece [10] proposed human detection method using fish-eye image to detect ellipses from the subtraction images of fish-eye pictures as human area. However, in a more crowded situation and when sudden illumination changes occur, their method shows a clear increase in false alarm rate.

In order to improve the efficiency of human detection on fish-eye images even in crowded indoor environments, we propose a human detection method. The rotations and sizes of the human regions on the fish-eye image change based on the locations of humans in the surveillance area. We propose a method to normalize these regions. Because a fish-eye lens camera is set on top of the surveillance area, the shapes of humans are changed based on their locations in the surveillance area. In this letter, we create three types of classifiers to detect humans in any part of the surveillance area; the most adaptive classifier for each human is chosen automatically from several classifiers. Moreover, we propose a method to minimize the occlusion effects. We infer the possible occlusion region in each human candidate region based on its location on the fish-eye image. Once the occluded regions are detected, the occlusion effects can be minimized by adjusting the threshold of the classifier.

Unlike other systems such as those proposed in Refs. [11,12], the human regions in our proposed method are detected initially from the fish-eye image, and only the human regions are corrected afterwards. In other systems, the entire input fish-eye images are corrected first and then the human regions are detected from the corrected images. Using our system, the processing efficiency can be improved and the processing time can be significantly reduced.

The system is designed as illustrated in Fig. 1, wherein the fish-eye lens camera is set on top of the surveillance area. The input image of the fish-eye lens camera is illustrated in Fig. 2, with the background image illustrated in Fig. 2(a) and the input image illustrated in Fig. 2(b).

The edges of the background and input images are extracted using Sobel operator[13] as illustrated in Figs. 3(a) and (b). In addition, the subtraction image between the input edge image and the background edge image is computed, as illustrated in Fig. 3(c). As shown in Fig. 3, all the head edges look like ellipses; thus, an efficient ellipse detection method[14] is adopted to extract the el-

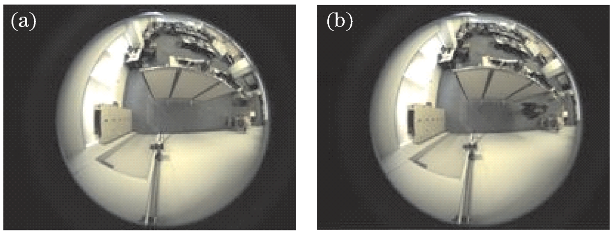

Fig. 1. Image taken by the proposed surveillance system. (a) Background image; (b) input image.

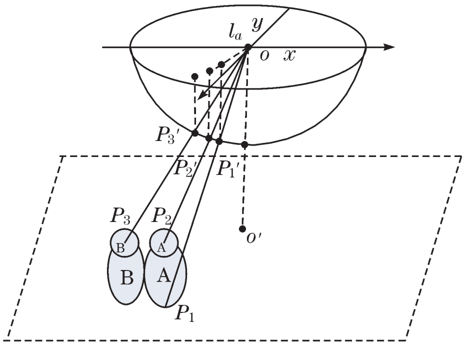

Fig. 2. Proposed surveillance system.

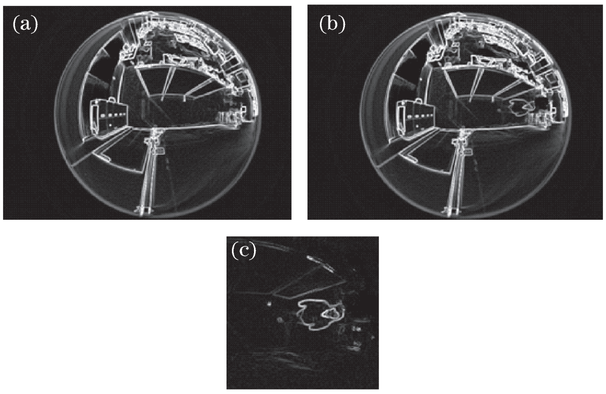

Fig. 3. Examples of edge images. (a) Background edge image; (b) input edge image; (c) subtraction image.

lipses from the edge image as head candidates. The proposed method is presented as follows.

For each pair of pixels, (x_1, y_1) and (x_2, y_2), the following five parameters of an ellipse can be calculated:

$$x_0 = (x_1 + x_2)/2, \tag{1}$$

$$y_0 = (y_1 + y_2)/2, \tag{2}$$

$$a = [(x_2 - x_1)^2 + (y_2 - y_1)^2]^{1/2}/2, \tag{3}$$

$$\alpha = a\tan[(y_2 - y_1)/(x_2 - x_1)], \tag{4}$$

$$b^2 = (a^2 d^2 \sin^2 \tau)/(a^2 - d^2 \cos \tau), \tag{5}$$

$$\cos \tau = (a^2 + d^2 - f^2)/2ad, \tag{6}$$

where (x_0, y_0) is the center of the assumed ellipse, a is the half-length of the major axis, α is the orientation of the ellipse, f is the focus of the ellipse, b is the half-length of the minor axis, and d is the distance between (x, y) and (x_0, y_0). A one-dimensional accumulator array is then used to vote on the half-length of the minor axis; if the votes reach a threshold, an ellipse is found and we yield the parameters for the detected ellipse and remove all pixels on that ellipse from the image.

The results of the extracted head candidates are illustrated in Fig. 4. The following process will be executed for each head candidate.

Based on the location of the head candidate, the method introduced in Ref. [15] was adopted in this letter in order to determine the size of the human candidates in different locations. Considering a cube whose size is bigger than a normal man standing on the floor in the real world, the projection of the cube can be considered as the human candidate region when the coordinate of the upper cube's projection is near the head candidate.

As shown in Fig. 1, points P_1', P_2', and P_3' in the fish-eye image are the projections of points P_1, P_2, and P_3 in real word, respectively. The projections of humans A and B are illustrated in Fig. 5. All the projections of humans seem to stand on the line l_a between the center of the fish-eye image (O) and their head candidate center (P_2' and P_3'); the feet of the human (P_1') are always closer to the center of the fish-eye image (O) than the human's head (P_2'). The angles α between the vertical line and the line from the center of the hand candidates are computed, and all the human candidate regions are rotated using

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{7}$$

where (x, y) is the coordinate of the original image and (x', y') is the coordinate of the rotated image. The results of the normalized human candidate regions are illustrated in Fig. 6.

As shown in Fig. 5, when humans stay at different locations, their shapes will change, making it impossible to detect all of them using the same detector. In this letter, we created three types of classifiers to categorize human and non human. Selecting the most adaptive classifier is thus an important issue. The shapes of humans change based on their distances from the head candidate center to the center of the image. We constructed three classifiers using different training images, with the most adaptive classifier selected using the following rules: If $d_\mathrm{i} \leq \theta_1$, classifier 1 is activated; if $\theta_1 \leq d_\mathrm{i} \leq \theta_2$, classifier 2 is activated; if $d_\mathrm{i} \geq \theta_3$, classifier 3 is activated. Here d_i is the distance between the center of the fish-eye image and the head candidate center; θ_1, θ_2, and θ_3 are some constants related to the threshold values.

We propose a method for selecting the values of θ_1, θ_2, and θ_3. We consider a cube whose size is bigger than a normal man standing on the floor in the real world and whose height is three times longer than its width.

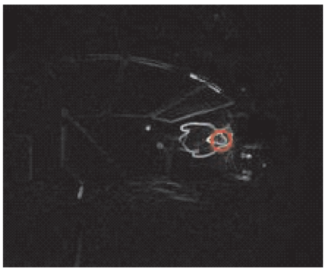

Fig. 4. Results of extracted human head candidates.

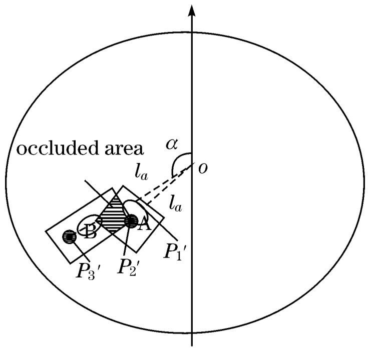

Fig. 5. Projections of humans standing on the floor.

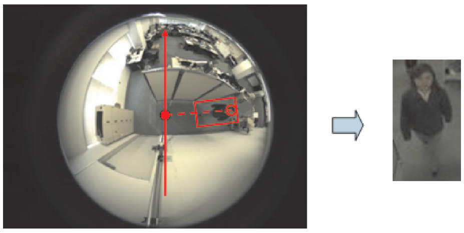

Fig. 6. Results of normalized human candidate regions.

Assuming that the cube moves from a distance and its projection is observed, when the height of the rectangle becomes 1.5 times longer than its width, θ_2 equals the distance between the upper point of the cube's projection and the center of the image. In the same way, when the height of the rectangle becomes equal to its width, θ_1 equals the distance between the upper point of the cube's projection and the center of the image.

In order to achieve accurate human detection, we adopt the histograms of oriented gradient (HOG) descriptors and use 13 cascade AdaBoost classifiers[7,8] to construct the three classifiers.

Since the human detection approach controls partial occlusions poorly, its accuracy may decrease in crowded conditions. Hence, we propose a method to handle occlusion effects.

As shown in Fig. 1, human A is closer to the center of the surveillance area (O') than human B in the real world, thus the projection of human A on the fish-eye image is also closer to the center of the fish-eye image (O) than human B. Using the fish-eye image taken by our system, the location of each human in the real world can be estimated. In this system, the human candidates are listed in order of their distances from the center of the fish-eye image, and human detection will proceed following this sequence.

If a human candidate region is evaluated as a human, this rectangle area is labeled as human A, as illustrated in Fig. 5. When a portion of the human candidate region is occluded by a human who has been previously detected, the occluded area (shown as the striped quadrilateral area in Fig. 5) and the occluded ratio in this region are computed using

$$r_{\text{ocl}} = \frac{A_{\text{ocl}}}{A_{\text{all}}}, \tag{8}$$

where A_{ocl} is the occluded area and A_{all} is the human candidate area.

The ratio r_{ocl} is used to adjust the threshold of the classifier. In this system, we adjust the number of cascades to minimize the occlusion effects. Increasing the number of cascades causes the detection of humans to become more difficult; decreasing the number cascades makes humans easier to be detected. The number of cascades is adjusted based on r_{ocl} using the following rules: if $r_{\text{ocl}} < 0.3$, the number of cascades is adjusted to 11; if $0.3 \leq r_{\text{ocl}} < 0.5$, the number of cascades is adjusted to 9; if $r_{\text{ocl}} > 0.5$, the human is considered to be the same person as the human in the front.

Fish-eye imaging system brings an inherent distortion in the image. Therefore, it is necessary to correct the human region to make it easier to understand. In our proposed system, we adopt the method introduced in Refs. [16,17] to correct the detected human region. The object plane shown in Fig. 7 is a typical region of interest; we aim to determine its mapping relationship onto the image plane to properly correct the perspective of the object. The image plane corresponds to the input fish-eye images. The direction-of-view vector, DOV(x,y,z), determines the zenith and azimuth angles for mapping the object plane onto the image plane, XY. The object plane is defined to be perpendicular to the vector $\overrightarrow{\text{DOV}}(x,y,z)$.

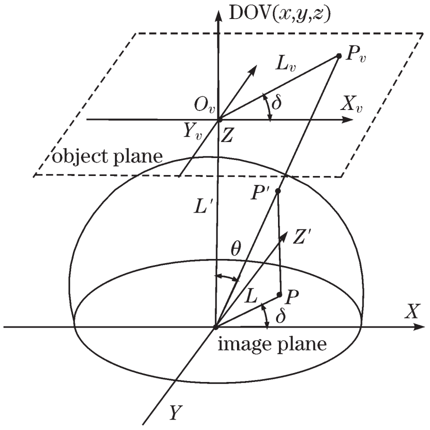

Fig. 7. Coordinate reference frame representation.

We define

$$x_v = l' \tan\left(\frac{\pi}{2} - \tan^{-1}\frac{\sqrt{R^2-(x^2+y^2)}}{\sqrt{x^2+y^2}}\right)\cos(\tan^{-1}\frac{y}{x}), \tag{9}$$

$$y_v = l' \tan\left(\frac{\pi}{2} - \tan^{-1}\frac{\sqrt{R^2-(x^2+y^2)}}{\sqrt{x^2+y^2}}\right)\sin(\tan^{-1}\frac{y}{x}), \tag{10}$$

where l' is the distance from the object plane to the image plane (see Fig. 7), and R is the radius of the fish-eye camera. Equations (9) and (10) provide a direct mapping from the XY image space to the X_vY_v space, providing the fundamental mathematical foundation for the omni-directional viewing system. By determining the desired zenith, azimuth, and object plane rotation angles and magnification, the locations of x and y in the input image can be established. Using Eqs. (3) and (4), the points in the object plane can then be computed, and the corrected human regions are achieved.

We collected images for training and video sequences of scenes for testing. The training data consisted of 2000 positive images and 2000 negative images for each classifier, whereas the test data consisted of 800 positive images and 800 negative images. The training data, as well as the test data, were captured on different places, and involved different people. All the test images were indoor scenes. The following conditions hold true: the maximum number of pedestrians in each scene is 10; 98 people are captured on the crowded conditions; the number of each scene is over 6; 110 people are captured when sudden illumination changes occur; 48 people are captured with the large cart. We compared five experiments to demonstrate the efficiency of our proposed method.

All the training data used for constructing classifier 1 follow this rule: the distances between the head candidate center of people and the center of the image are less than θ_1. Figure 8 shows the examples of detection results using only classifier 1; the rectangle shows the detected human region.

All the training data used for constructing classifier 2 follow this rule: the distances between the head candidate center of people and the center of the image are d_{i}, $\theta_1 \leq d_{\text{i}} \leq \theta_2$. Figure 9 shows the examples of detection

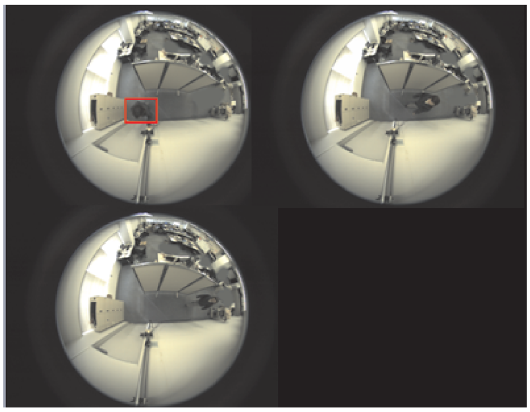

Fig. 8. Examples of human detection results using classifier 1.

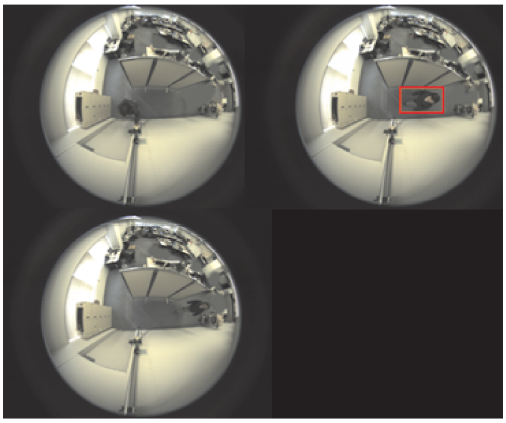

Fig. 9. Examples of human detection results using classifier 2.

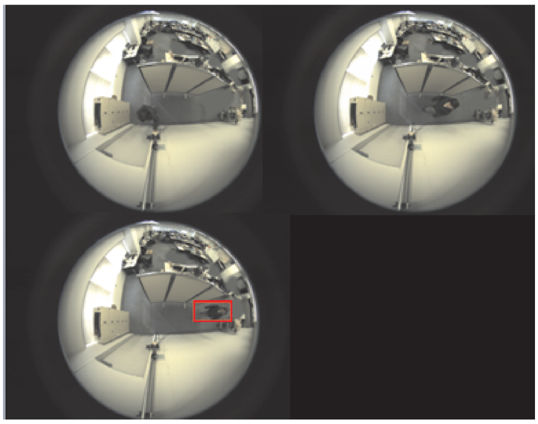

Fig. 10. Examples of human detection results using classifier 3.

results using only classifier 2; the rectangle shows the detected human region.

All the training data used for constructing classifier 3 follow this rule: the distances between the head candidate center of people and the center of the image are larger than θ_2. Figure 10 shows the examples of detection results using only classifier 3; the rectangles shows the detected human region.

The training data used for constructing the overall classifier include all the training data used for constructing classifiers 1, 2, and 3. The training data consisted of 6000 positive images and 6000 negative images. Figures 11−13 show the examples of detection results using the overall classifier; the rectangles are the detected human regions. As shown in the figures, too many false alarms appear using the overall classifier.

In Ref. [10], the authors proposed a method to detect

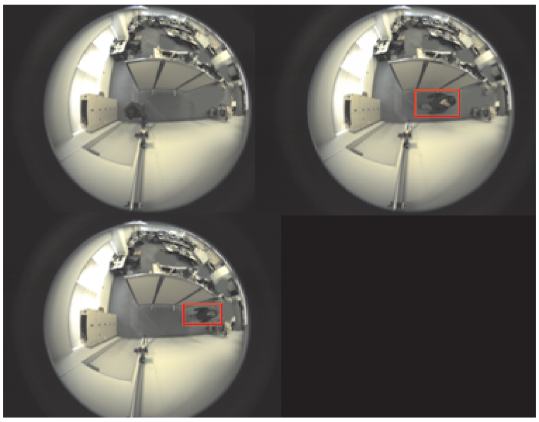

Fig. 11. Examples of human detection results using the overall classifier.

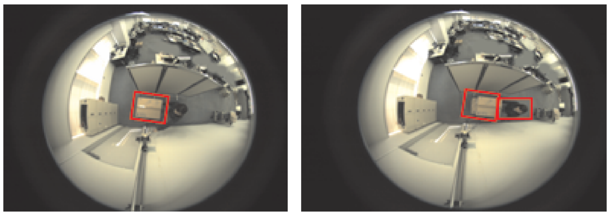

Fig. 12. Examples of human detection results using the overall classifier (human with a large cart).

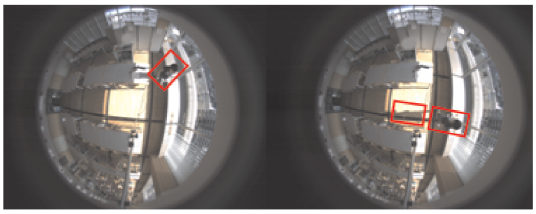

Fig. 13. Examples of human detection results using the overall classifier (wherein sudden illumination changes occur).

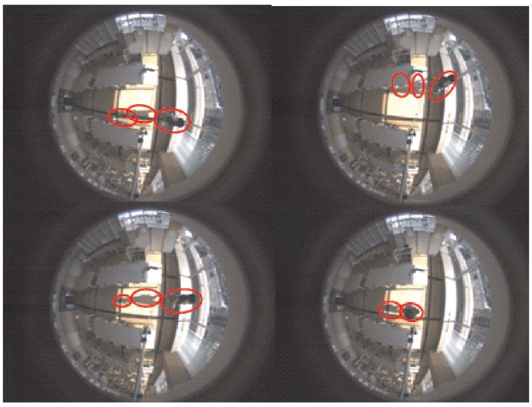

Fig. 14. Examples of human detection results using the ellipse detection method.

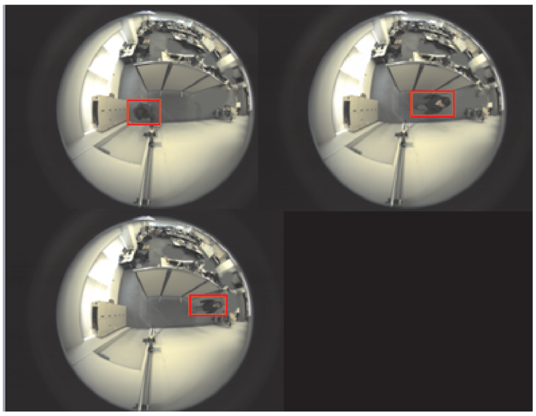

Fig. 15. Examples of human detection results using the proposed method.

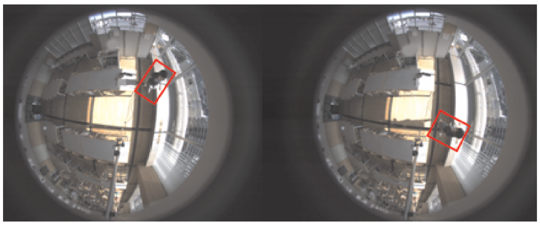

Fig. 16. Examples of human detection results using the proposed method (human with a large cart).

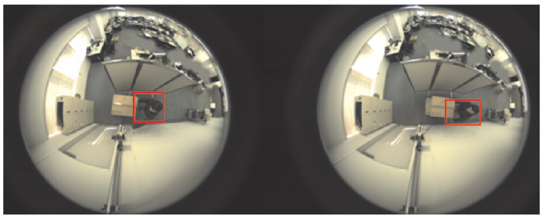

Fig. 17. Examples of human detection results using the proposed method (wherein sudden illumination changes occur).

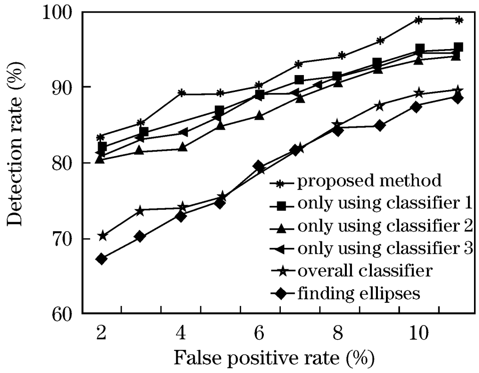

Fig. 18. Human detection results.

ellipse regions as human regions. The experiment for comparison was carried out using the same testing images that were used in our proposed system. Experimental results show that the performance of their method is very poor when sudden illumination changes occur. Figure 14 shows the examples of detection results using the ellipse detection method.

As shown in Figs. 15−17, the proposed method may successfully detect humans in fish-eye images. The experimental results show that the performance of the proposed method is better than any other methods.

A receiver operating characteristic (ROC) curve, which plots the detection rate versus the false positive rate, shows the experimental results (Fig. 18). With a false negative rate of 10%, our method has a false positive rate that is 9.75% lower than the system using only classifier 1 for classification, 5.25% lower than the system using only classifier 2, 4.5% lower than the system using only classifier 3, 4% lower than the system using the overall classifier, and 11.25% lower than the system that finds ellipses as humans. These results indicate that our proposed method has better accuracy compared with other methods.

In conclusion, we present an automatic surveillance system using fish-eye lens camera. We propose a human detection method using fish-eye lens camera, which achieves wide-area automatic surveillance without a dead angle using only one camera. The experimental results demonstrate the efficiency of our algorithm.

References

1. T. Zhao, R. Nevatia, and B. Wu, IEEE Trans. Pattern Anal. Machine Intell. **30,** 1198 (2008).
2. N. Dalal and B. Triggs, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2005).
3. P. Felzenszwalb, D. McAllester, and D. Ramanan, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2008).
4. C. H. Lampert, M. B. Blaschko, and T. Hofmann, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2008).
5. X. Wang, T. X. Han, and S. Yan, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2009).
6. P. Viola and M. Jones, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2001).
7. Y. Yamauchi, T. Yamashita, and H. Fujiyoshi, IEICE Trans. Inform. Syst. (in Japanese) **J92-D,** 1125 (2009).
8. X. Yuan, X. Wei, and Y. Song, "Pedestrian detection for counting applications using a top-view camera" IEICE Trans. Inform. Syst. (to be published).
9. G. Jia, X. Wang, and S. Zhang, Acta Opt. Sin. (in Chinese) **29,** 659 (2009).
10. Y. Kubo, T. Kitaguchi, and J. Yamaguchi, in *Proceedings of SICE Annual Conference 2007* (2007).
11. W. Kim and C. Kim, in *Proceedings of IEEE International Symposium on Circuits and Systems* (2009).
12. S. Zimmermann and D. Kuban, in *Proceedings of IEEE/AIA Digital Avionics Systems Conference* 523 (1992).
13. T. Zhang, Y. Lu, and X. Zhang, Acta Opt. Sin. (in Chinese) **29,** 180 (2009).
14. Y. Xie and Q. Ji, in *Proceedings of 16th International Conference on Pattern Recognition* (2002).
15. H. Arai, I. Miyagawa, H. Koike, and M. Haseyama, in *Proceedings of International Conference on Pattern Recognition* (2008).
16. G.-I. Kweon, Y.-H. Choi, and M. Laikin, J. Opt. Soc. Korea **12,** 79 (2008).
17. J. Wu, K. Yang, Q. Xiang, and N. Zhang, Chin. Opt. Lett. **7,** 142 (2009).