A fast moving object detection method based on 2D laser scanner and infrared camera

Lina ZENG[1], Meng DING[*1], Tianci ZHANG[1], Zejun SUN[1]
1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R.China

ABSTRACT

Moving object detection is a major research direction of video surveillance systems. This paper proposes a novel approach for moving object detection by fusing information from the laser scanner and infrared camera. First, in accordance with the feature of laser scanner data, we apply robust principal component analysis (RPCA) to studying moving object detection. Then the depth and angle information of moving objects is mapped to the infrared image pixels so as to obtain the regions of interest (ROI). Finally, moving objects can be recognized by making investigation of the ROI. Experimental results show that this method has good real-time performance and accuracy.

Key words: 2D laser scanner, infrared camera, moving object detection, robust principal component analysis, regions of interest

1. INTRODUCTION

The technology of intelligent video surveillance that can replace human operators to monitor the areas under surveillance is becoming more and more important in many fields of modern society. Moving object detection (MOD) is one of the core functions of video surveillance systems. Accurate MOD is the technical foundation of object recognition and object tracking.

Background subtraction based on background modeling [1, 2] is one of the most common approaches to detect moving objects, but there are many challenges of visual-based MOD such as bad performance in real-time applications and illumination changes. It is difficult to simultaneously improve the real-time performance and the detection accuracy by merely designing a new and complex algorithm. And a reasonable approach is to fuse information from multiple sources, we choose the fusion of laser scanner and infrared camera information in this paper. There are two reasons for that: Laser scanner has an advantage in the real-time performance; visible images cannot capture targets effectively in low visibility conditions compared with infrared images [3]. We narrow the searching range of infrared images down to the regions of interest (ROI) by guiding of the angle and distance information of the moving objects obtained by a laser scanner so that we can make fast recognition of moving objects.

Wang et al. [4] proposed a multi-hypothesis tracking data association algorithm to detect moving objects. But it is hard to detect temporarily static objects, and the computational complexity of tracking increases exponentially. Azim et al. [5] used Bayesian inference to update the grid map, and took a clamping update policy that could control occupation probability of the grid map in a certain range $(\varepsilon, 1-\varepsilon)$. This article detected moving objects by comparing with the grid map at the current moment and the global map at the former moment. However, the algorithm has the defects of calculation complexity and there is a large amount of noise in the grid map.

In our work, we follow another direction which uses robust principal component analysis (RPCA) [6, 7] to get foreground matrix by decomposing the observation matrix constructed from continuous multi-frame laser scanner data. Because there is intensity noise in the foreground, a window filtering algorithm is designed so that we can accurately obtain the depth and angle information of moving objects. The mapping relation between the laser scanner polar coordinates and the infrared image pixel coordinates is established under the premise that the installations of both laser scanner and infrared camera are fixed. Thus, the depth and angle information of moving objects obtained can be mapped to the corresponding infrared image pixels in order to define the ROI. Finally, on the basis of histogram of oriented gradient (HOG) [8, 9] feature extraction, the classifier is trained off line using support vector machine (SVM) to identify objects in the obtained ROI. Fig.1 depicts the proposed framework.

* Corresponding author. E-mail: nuaa_dm@hotmail.com

AOPC 2015: Image Processing and Analysis, edited by Chunhua Shen, Weiping Yang, Honghai Liu,
Proc. of SPIE Vol. 9675, 96752K · © 2015 SPIE · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2202698

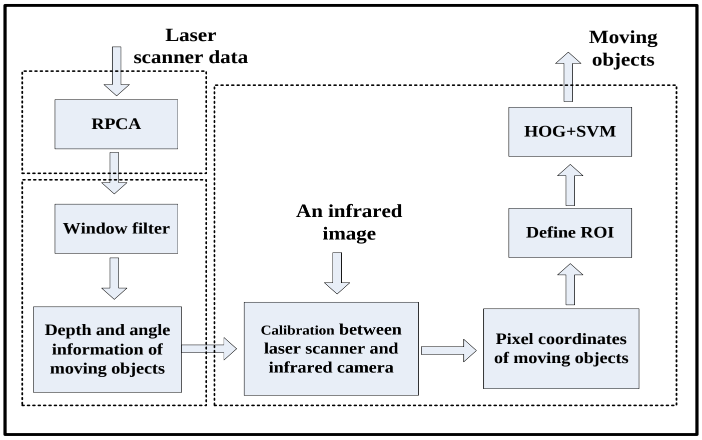


Fig.1 The proposed framework

2. MOVING OBJECT DETECTION

2.1 Moving object detection based on RPCA

Let $\{d_1, d_2, \ldots, d_k\}$ be the depth information sequence of k continuous frames from a stationary 2D laser scanner; $d_i = \{d_{n,i} \,|\, d_{n,i} = f(\theta_n), n = 1, \ldots, m\} (i = 1, \cdots, k)$, where $d_{n,i}$ represents measurement distance of the n-th scan point, θ_n is the azimuth of the n-th scan point, m denotes the number of scan points of the laser scanner.

In order to detect moving objects from continuous multi-frame laser scanner data via RPCA, it is necessary to determine the relationships among the laser scanner data, background and foreground (shown in Fig. 2). The background is a single frame of laser scanner containing no depth information of objects of interest. By subtracting depth values from a frame containing objects of interest, any non-zero values are classified as the foreground.

The purpose of MOD is to differentiate moving foreground objects from the background. This means the laser scanner data d_i at the i-th frame can be decomposed into a background term a_i and foreground term e_i, that is $d_i = a_i + e_i$. If each frame laser scanner data is aligned as a column vector $\mathbf{D} = [d_1, \ldots, d_k] \in R^{m \times k}$, then the decomposition of the laser scanner data sequence can be represented as **D**=**A** + **E**, where $\mathbf{A} = [a_1, \ldots, a_k] \in R^{m \times k}$ is the background term and $\mathbf{E} = [e_1, \ldots, e_k] \in R^{m \times k}$ is the moving object term. In theory, the linear correlations of background terms $\{a_i\}$ are strong. Therefore, the background term **A** should essentially be a low rank matrix. The number of points corresponding to the moving object is usually small compared to the total number of scan points, so the foreground moving object can be treated as sparse. Recover of the low rank matrix **A** and sparse matrix **E** via RPCA can be described as the following optimization process:

$$\min_{\mathbf{A},\mathbf{E}} \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \quad s.t.\, \mathbf{D} = \mathbf{A} + \mathbf{E} \tag{1}$$

Where, $\|\mathbf{A}\|_* = \sum_{\mathbf{i}=1}^{\mathbf{m}} \delta_{\mathbf{i}}(\mathbf{A})$, $\delta_{\mathbf{i}}(\mathbf{A})$ represents the i-th singular value of the matrix $\mathbf{A}$; $\|\mathbf{E}\|_1 = \sum_{\mathbf{ij}} |\mathbf{E_{ij}}|$, $\mathbf{E_{ij}}$ denotes the values of matrix $\mathbf{E}$. λ is a positive weighting parameter.

As a sparse matrix, the entries of matrix $\mathbf{E}$ should be 0 except those representing moving objects. In fact, because of the noise disturbance and iteration calculation of RPCA, some noise terms occur in foreground matrix which are not equal to 0. Therefore, in order to detect moving objects accurately, we should further process the foreground matrix representing.

2.2 Window filtering

Due to the presence of noise, it is necessary to filter the sparse foreground matrix $\mathbf{E}$. Generally speaking, moving objects related terms have the characteristic that the values of adjacent scan angle have strong correlations in each column of $\mathbf{E}$. According to this, we design a dynamic adaptive window filtering algorithm to eliminate noise.

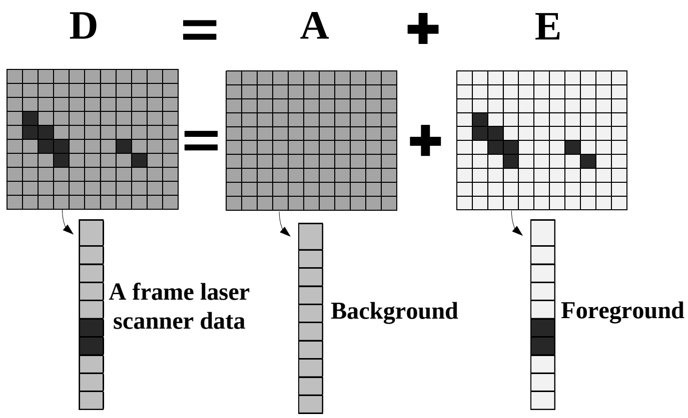


Fig.2 Sketch map of the relationship between the observation matrix **D**, low rank matrix **A** and sparse matrix

The first step is to transform the non-zero entries of foreground matrix **E** obtained directly by RPCA to the corresponding depth information, getting a new foreground matrix $\mathbf{E_N}$. Then, we fetch data which can be expressed by $d_{i,j}$ from the i-th row j-th column of matrix $\mathbf{E_N}$ and assume that $d_{i,j}$ is not equal to 0. The adaptive threshold value $t \in N^*$ (shown in table.1) should be selected based on $d_{i,j}$ and t is determined by depth information, angular resolution of the laser scanner and the minimum width of moving objects (assumed 0.5m).

Table 1.the value of t within different range of distance

Depth(m)	(0,5]	(5,10]	(10,20]	(20,30]
t	23	11	6	4

Then the data window is established as follows:

$$P = \left\{ d_{i,j}, d_{i+1,j}, d_{i+2,j}, \ldots, d_{i+t-1,j} \middle| i+t-1 \le m \right\} \tag{2}$$

Denote the number of the non-zero entries among P as q. If $q < t$, then set $d_{i,j}$ to zero; otherwise keep it unchanged. When all non-zero entries have been examined, we can get the new matrix $\mathbf{E_N}'$ in which the non-zero entries are depth values of moving objects.

2.3 Calibration and recognition

The approach makes use of the reliability of the laser scanner to define ROI in the infrared images. Only depth and angle information of the moving objects detected by the laser scanner are provided to the infrared image so as to reduce the search area of the infrared image of the recognition algorithms. Thus computation cost is reduced.

Coordinate conversion between infrared image coordinate system and infrared camera coordinate system can be performed using pin-hole model (3) and the mapping relation between laser scanner coordinate system and infrared camera coordinate system is established by accurate extrinsic calibration (4):

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{3}$$

Where λ is the scale factor, u_0 and v_0 are the center coordinates of the infrared camera coordinate system in pixels. (u,v) are the coordinates in the infrared camera coordinate system in pixels. x, y and z are the Cartesian coordinates of the infrared camera. And f is the focal length.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T \tag{4}$$

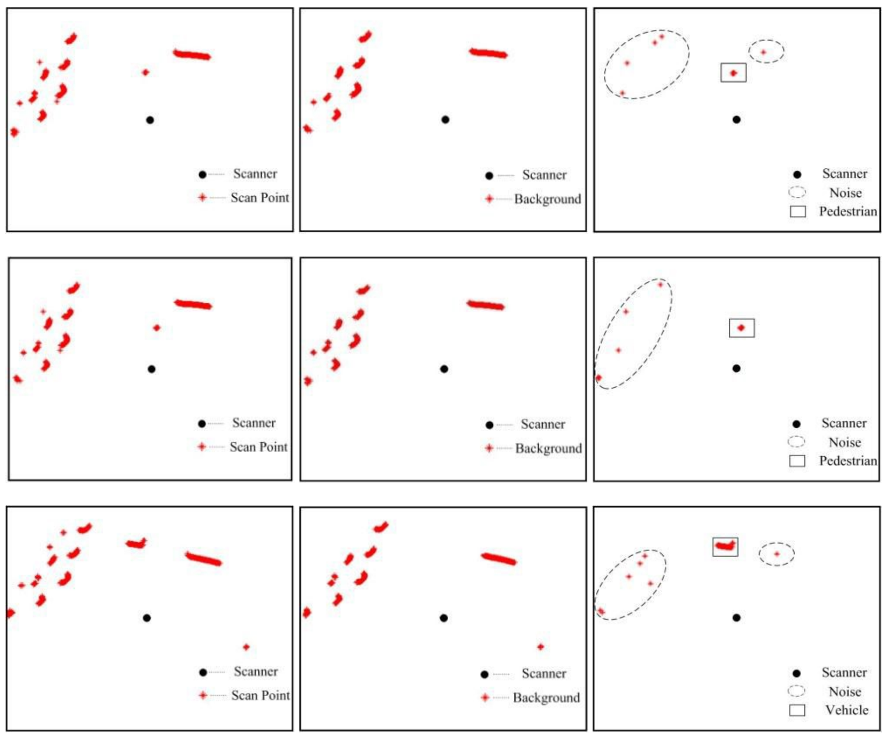
Where R is the rotation matrix, T is the translation vector, (x_1, y_1, z_1) are the coordinates in the laser scanner Cartesian coordinate system which are converted from polar coordinate system and (x, y, z) are the infrared camera coordinates.

Since the foreground of the single frame is from each column of matrix $\mathbf{E_N}'$, we can get depth information of moving objects of each frame. Then the depth and angle information of moving objects can be mapped to the corresponding infrared image in order to get ROI, which contains complete information of moving objects.

In this paper, we focus on the detection of vehicles and pedestrians. First, the machine learning method is used to extract the histogram of oriented gradient (HOG) feature of the pedestrian or vehicle. Then the classifier is obtained via training a support vector machine (SVM), which is used to recognize pedestrians or vehicles.

3. EXPERIMENTS AND DISCUSSIONS

This section demonstrates the performance of the proposed methodology with experiments in a real road surveillance scenario. In the experiments, 5000 continuous frames laser scanner data are selected for testing. The results of the 1386-th frame, 1483-th frame, and 2759-th frame are illustrated in Fig.3 and Fig.4.

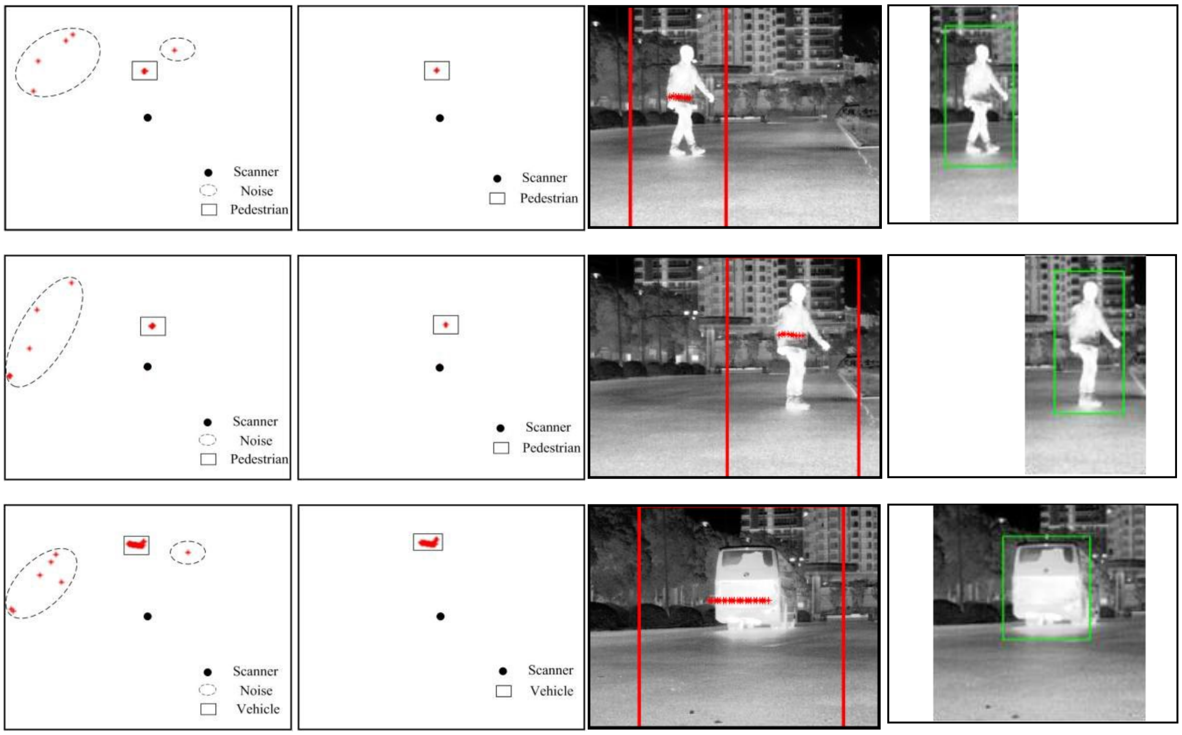


(a) Original data (b) Background (c) Foreground

Fig.3 Background and foreground segmentation via RPCA

We take the 5000 frames laser scanner data to construct observation matrix **D** and then recover matrix **A** and **E** via RPCA. The 1386-th column of **D**, **A** and **E** are shown in the first row in Fig.3. The second and third rows correspond to the 1483-th column and 2759-th column respectively. From Fig.3, we can see that using the original data of the laser scanner that contain moving objects (shown in Fig.3(a)), the background (shown in Fig.3(b)) and foreground (shown in Fig.3(c)) can be segmented via RPCA.

As we can see in Fig.3(c) or Fig.4 (a), the foreground matrix generated by RPCA is composed of the moving objects and the intensity noises. The noises are then eliminated by window filtering as shown in Fig.4 (b) so that the efficiency and accuracy of subsequent processing can be greatly improved. Then the object points are projected onto the infrared image based on the projective transformation matrix and the ROI is labeled using red lines (see Fig.4 (c)).The final result of recognition acquired by making investigation within the ROI is demonstrated in Fig.4 (d).The pedestrian or vehicle on the road is clearly bounded out. Additional experimental results also demonstrate that if we use the original infrared image to recognize the moving object directly, the time will increase at least 5 times when there is only one moving object. Therefore, the proposed approach can greatly improve the real-time performance.



(a) Foreground (b) Foreground via window filter (c)Regions of interest (ROI) (d) Detections based on ROI

Fig.4 Results of moving object detection

4. CONCLUSION

This paper presents a fusion method which combines the information from the infrared camera and the laser scanner. Limitation of laser scanner that cannot be used for precise recognition of object type is overcome thanks to the use of fusion method. Experimental results demonstrate that the proposed method can provide reliable MOD via RPCA based on laser scanner data and help to improve the real time performance.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.61203170) and Special Foundation of China Postdoctoral Science (No. 2013T60539).

REFERENCES

[1] Haritaoglu, I., Harwood, D. and Davis, L., “real-time surveillance of people and their activities,” IEEE Trans Pattern Analysis and Machine Intelligence, Papers 22(8), 809-830 (2000).

[2] Moeslund, T., Hilton, A. and Kruger V., “A Survey of Advances in Vision-Based Human Motion Capture and Analysis,” Computer Vision and Image Understanding, Papers 104(2-3), 90-126 (2006).

[3] Ding, M., Wei, L., Wang, B.F., “Research on fusion method for infrared and visible images via compressive sensing,” Infrared Physics & Technology, Papers 57(2), 56-67 (2013).

[4] Wang, C.C., Thorpe, C. and Thrun, S., “Online Simultaneous Localization and Mapping with Detection and Tracking of Moving Objects: Theory and Results from a Ground Vehicle in Crowded Urban Areas,” Proceedings of the 2003 IEEE International Conference on Robotics and Automation, 842-849 (2003).

[5] Azim, A., Aycard, O., “Detection, classification and tracking of moving objects in a 3D environment” Proc. of the Intelligent Vehicle Symposium, 802-807 (2012).

[6] Wright, J., Peng, Y.G., Ma, Y. and Rao, S., “Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization‖,” Proceedings of the Conference on Neural Information Processing Systems (NIPS), (2009).

[7] Ganesh, A., Lin, Z.C. and Wright, J.,”Fast algorithms for recovering a corrupted low-rank matrix,” IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 213-216 (2009).

[8] Dalal, N. and Triggs, B. “Histograms of oriented gradients for human detection,” Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 886-893 (2005).

[9] Wang, Z.R., Jia, Y.L., Huang, H. and Tang, S.M. “Pedestrian detection using boosted HOG features,” Proceedings of the 11th International Conference on Intelligent Transportation Systems, 1155- 1160 (2008).