

Moving Object Detection Using a Stereo Camera Mounted on a Moving Platform

Tetsuro TODA *, Gakuto MASUYAMA **, and Kazunori UMEDA **

Abstract: In this paper, we propose a method for detecting moving objects using a moving stereo camera. First, the camera motion parameters are estimated by using optical flow with a stereo camera. Second, the optical flow occurring in the background is removed. Finally, moving objects are detected individually by labeling the remaining optical flow. The proposed method is evaluated through experiments using two pedestrians in an indoor environment.

Key Words: moving object detection, stereo camera, optical flow.

1. Introduction

To achieve the automatic operation of mobile robots and automobile, the detection of moving objects has become increasingly important. This paper presents a system of detecting moving objects using images obtained from a camera mounted on an object such as a car and a mobile robot.

In images obtained from a moving camera, both moving objects and the background could appear to move as shown in Fig. 1. Therefore, it is not plausible to apply a simple background subtraction method to detect moving objects in the scene. Some methods for applying background subtraction to the detection of moving objects from a moving camera were proposed. An algorithm using a multiple background model based background subtraction system was presented in [1]. Zamalieva *et al.* [2] proposed a method to apply background subtraction after estimating the geometry of the scene from several consecutive frames using the temporal fundamental matrix. On the other hand, there are some methods that do not use background subtraction. Hu *et al.* [3] proposed a method of detecting a moving object from a car-mounted camera. Rodriguez *et al.* [4] used a parallel tracking and mapping (PTAM) [5] algorithm to estimate the camera's motion parameters. Moving objects can be detected by comparing the optical flow obtained from the images and artificial optical flow obtained from motion parameters. These two methods [3],[4] are designed for cameras mounted on cars and unmanned aerial vehicles (UAVs) that have limited camera motion and orientation. As a different approach, Takeda *et al.* [6] proposed a method using focus of expansion (FoE) of the optical flow. Furthermore, a method of detecting moving objects from FoE using normal flow which can be calculated directly from the image sequence data was proposed in [7]. These approaches are not limited to the camera's movement and orientation. However, they can-

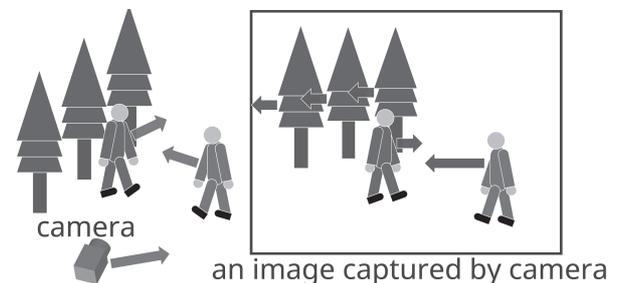


Fig. 1 An illustrative image captured from a moving camera. Arrows of left image represent movements of the camera and pedestrians; arrows of right image represent movements of objects in the image.

not detect moving objects when the camera and the moving objects are moving in the same direction. As a different approach, Mochizuki *et al.* [8] proposed a method using template matching based on background elimination. Another method based on stereo cameras and a neural network was proposed by Zhao *et al.* [9]. Methods for advanced driver assistance system (ADAS) using adaptive boosting (AdaBoost) for classification to detect vehicles are presented in [10],[11]. These methods using template matching or a neural network can only detect a prespecified target because the template of the target must be prepared.

We propose a method that takes a different approach using a stereo camera to detect moving objects based upon three-dimensional (3D) coordinates. The method is free from the limitations of camera's motion and direction. The flow of the system is described in Fig. 2, which is based on [12].

First, feature points are detected from two related frames of the current image and the previous image. The optical flow is acquired by matching feature points between the two images. The optical flow is then classified into three types: moving object flow, background flow and other than optical flow due to false matching. Our purpose is to remove all optical flow except for moving object flow.

Second, camera motion parameters are estimated using 3D information of each feature point. The calculation method of camera motion is similar to the one proposed by Talukder *et al.* [13].

Third, optical flow representing the apparent movement of

* Graduate School of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan

** Faculty of Science and Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo, Japan
E-mail: toda@sensor.mech.chuo-u.ac.jp,
masuyama@mech.chuo-u.ac.jp,
umeda@mech.chuo-u.ac.jp

(Received January 20, 2017)

(Revised April 14, 2017)

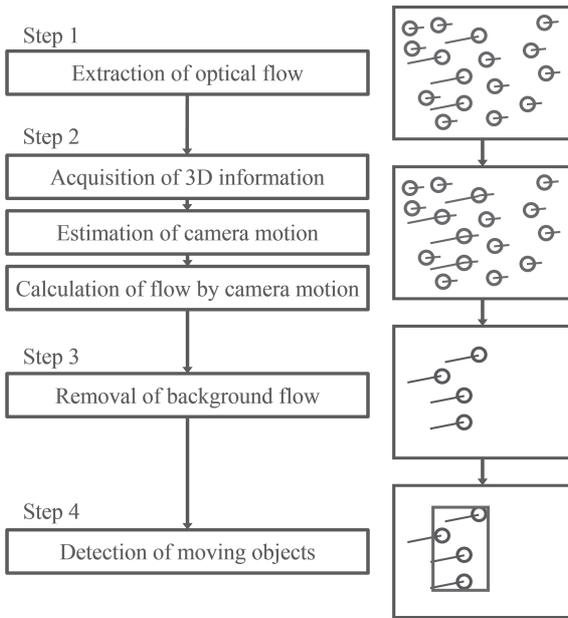


Fig. 2 Flow of the proposed method.

the background is estimated. Hereinafter, the estimated optical flow is denoted by “estimated flow.” The background flow is removed by comparing the estimated flow and the optical flow obtained from the image.

Finally, the moving object is detected by extracting the moving object’s flow from the remaining optical flow.

The rest of the paper is organized as follows. Section 2 explains the algorithm of the method. In Section 3, we conduct experiments on detecting moving objects to show the validity of the proposed method. The performance of the proposed method is compared with that of [12]. Finally, our conclusions and future work are shown in Section 4.

2. Detection of Moving Objects

2.1 Overview of Approach

For steps 1 to 3 of Fig. 2, the processing is performed in the same way as in [12]. In the first step, feature points are extracted. The optical flow is then extracted by matching them between two consecutive frames. Accelerated KAZE (AKAZE) [14] is used to extract feature points. In the second step, the motion parameters of the camera are estimated. The 3D coordinates of feature points necessary for camera motion estimation are acquired from a stereo camera. Thus, we only have to calculate the 3D coordinates of the feature points; the disparity of an entire image is not required. Also, outliers included in the result of feature-point matching are removed by random sample consensus (RANSAC) [15]. Furthermore, estimated flow representing the apparent motion of the background is calculated using the equations in [13] in this step. In the third step, the background optical flow is removed by comparing the optical flow and the estimated flow. In the final step, moving objects are detected. First, the remaining feature points after removal of the background optical flow are labeled. Then, feature points of a moving object are extracted using a histogram on the Z axis with the feature points of the same label. Feature points associated with a bin taking a large number are used for detection.

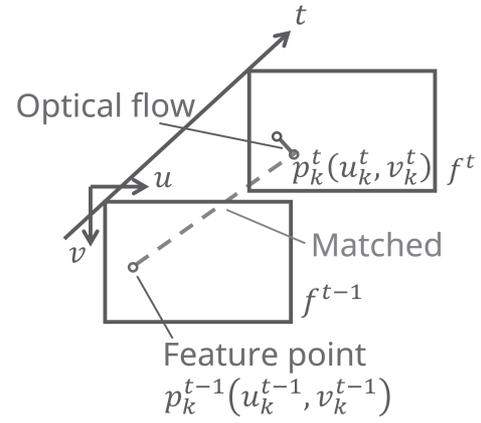


Fig. 3 Extracting optical flow.

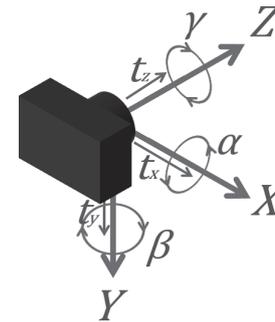


Fig. 4 Camera motion parameters.

2.2 Extracting Optical Flow

The extraction of optical flow is shown in Fig. 3. The current frame acquired from the camera is f^t , and the previous frame is f^{t-1} adjacent to f^t , where t is the time of the captured frame. Optical flow F_k is a movement vector from p_k^{t-1} to p_k^t , where p_k^{t-1} and p_k^t are the feature points in f^{t-1} and f^t , respectively, and $k = 1, \dots, n$ (n is the number of matched feature points). The descriptor of the AKAZE algorithm is robust to scale change, rotation and blur. Therefore, the AKAZE algorithm is suited to obtaining optical flow from the moving camera.

2.3 Estimating Camera Motion Parameters

Figure 3 shows the camera motion parameters. The translation vector t and the rotation matrix R are represented as

$$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{1}$$

$$R = \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix}, \tag{2}$$

where R is a linearized rotation matrix approximated by the rotation angle of the camera, which is assumed to be negligibly small, and $\alpha, \beta,$ and γ are the rotation angles around the $X, Y,$ and Z axes respectively, as shown in Fig. 4.

The 3D coordinates of the feature points p_k^t and p_k^{t-1} are $P_k^t = [X_k^t \ Y_k^t \ Z_k^t]^T$ and $P_{k-1}^t = [X_{k-1}^t \ Y_{k-1}^t \ Z_{k-1}^t]^T$, respectively. Then, the relationship of P_k^t and P_{k-1}^t is described as follows:

$$P_k^{t-1} = RP_k^t + t. \tag{3}$$

We can rewrite Eq. (3) to isolate the unknown terms as

$$\mathbf{P}_k^{t-1} - \mathbf{P}_k^t = \mathbf{R}' \mathbf{P}_k^t + \mathbf{t}. \quad (4)$$

where \mathbf{R}' is represented as

$$\mathbf{R}' = \begin{bmatrix} 0 & -\gamma & \beta \\ \gamma & 0 & -\alpha \\ -\beta & \alpha & 0 \end{bmatrix}. \quad (5)$$

Then, the relationship between \mathbf{P}_k^t , \mathbf{P}_k^{t-1} , and the parameters is described as

$$\mathbf{A}_k \boldsymbol{\xi} = \mathbf{b}_k, \quad (6)$$

where

$$\mathbf{A}_k = \begin{bmatrix} 0 & Z_k^t & -Y_k^t & 1 & 0 & 0 \\ -Z_k^t & 0 & -X_k^t & 0 & 1 & 0 \\ Y_k^t & -X_k^t & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (7)$$

$$\boldsymbol{\xi} = \begin{bmatrix} \alpha & \beta & \gamma & t_x & t_y & t_z \end{bmatrix}^T, \quad (8)$$

$$\mathbf{b}_k = \mathbf{P}_k^{t-1} - \mathbf{P}_k^t. \quad (9)$$

The camera motion parameters are calculated by the linear least squares method as follows:

$$\boldsymbol{\xi} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix}^+ \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_n \end{bmatrix}, \quad (10)$$

where “+” represents the pseudoinverse and

$$\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix}^+ = \left(\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix}^T \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_n \end{bmatrix}. \quad (11)$$

The matching results contain outliers due to false matching or moving object flow; thus, outliers are removed by RANSAC [15].

2.4 Calculating the Estimated Flow

In the normalized image coordinate, the feature point $\mathbf{p}_k^t = [u_k^t \ v_k^t]^T$ in the image coordinate is represented by $\mathbf{q}_k^t = [x_k^t \ y_k^t]^T$. The relationship between \mathbf{p}_k^t and \mathbf{q}_k^t is described as

$$x_k^t = \frac{\delta_u(u_k^t - c_u)}{f}, \quad (12)$$

$$y_k^t = \frac{\delta_v(v_k^t - c_v)}{f}, \quad (13)$$

where f is the focal length, δ_u and δ_v are the pixel size of the horizontal and vertical directions, respectively, and $[c_u \ c_v]^T$ is the image center. Then, the relationship between \mathbf{q}_k^t and \mathbf{P}_k^t is described as

$$x_k^t = \frac{X_k^t}{Z_k^t}, \quad (14)$$

$$y_k^t = \frac{Y_k^t}{Z_k^t}. \quad (15)$$

The estimated flow $\mathbf{F}'_k = [\dot{x}_k^t \ \dot{y}_k^t]^T$ is calculated as

$$\mathbf{F}'_k = \begin{bmatrix} \dot{x}_k^t \\ \dot{y}_k^t \end{bmatrix} = \begin{bmatrix} \frac{\partial x_k^t}{\partial t} \\ \frac{\partial y_k^t}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{X_k^{t-1}}{Z_k^t} - \frac{X_k^t Z_k^{t-1}}{Z_k^2} \\ \frac{Y_k^{t-1}}{Z_k^t} - \frac{Y_k^t Z_k^{t-1}}{Z_k^2} \end{bmatrix}, \quad (16)$$

where X_k^{t-1} , Y_k^{t-1} , and Z_k^{t-1} are represented from Eq. (3) as

$$X_k^{t-1} = X_k^t - \gamma Y_k^t + \beta Z_k^t + t_x, \quad (17)$$

$$Y_k^{t-1} = \gamma X_k^t - Y_k^t + \alpha Z_k^t + t_y, \quad (18)$$

$$Z_k^{t-1} = -\beta X_k^t - \alpha Y_k^t + Z_k^t + t_z. \quad (19)$$

Therefore, the estimated flow \mathbf{F}' is described as

$$\mathbf{F}'_k = \begin{bmatrix} -\gamma y_k^t + \beta + \frac{t_x}{Z_k^t} - x_k^t \left(-\beta x_k^t + \alpha y_k^t + \frac{t_z}{Z_k^t} \right) \\ -\gamma x_k^t + \alpha + \frac{t_y}{Z_k^t} - y_k^t \left(-\beta x_k^t + \alpha y_k^t + \frac{t_z}{Z_k^t} \right) \end{bmatrix}. \quad (20)$$

Then, a point $\mathbf{p}_k^{t'} = [u_k^{t'} \ v_k^{t'}]^T$ is estimated as

$$u_k^{t'} = u_k^t + \frac{f \dot{x}_k^t}{\delta_u}, \quad (21)$$

$$v_k^{t'} = v_k^t + \frac{f \dot{y}_k^t}{\delta_v}. \quad (22)$$

2.5 Removing the Background Flow

The Euclidean distance between \mathbf{p}_k^t and $\mathbf{p}_k^{t'}$ is represented as d_k . When d_k is lower than the threshold d_{th} , the optical flow \mathbf{F}_k is removed as background flow. Therefore, the proposed method can deal with any movement of camera and moving object, as long as the (relative) movement of the moving object makes some motion, i.e., optical flow on the image. When an object moves just towards or opposite to the camera, the movement does not produce optical flow (after compensation of the camera motion) and thus it cannot be detected. However, this can be avoided for a real application, because an object has some size and it is impossible that every feature point on the object moves towards the camera.

2.6 Detecting Moving Objects

The optical flow that remain after removing the background optical flow is expected to be optical flow extracted from the periphery of the moving objects. In order to individually detect moving objects, the feature points are labeled. First, the same labels are given to feature points whose distances are short on the image coordinates as shown in Fig. 5. Second, a histogram is made with respect to the Z axis of the 3D coordinates with the feature points of the same label as shown in Fig. 6. Circles in Fig. 6(a) represent the feature points. The feature points are those that remain after the removal of the background flow. Also, the rectangle represents the result of moving object detection. Figure 6(b) shows the histogram of Z in the scene, where Z is normalized from 0 to 1. Finally, a group of feature points belonging to a bin having a value larger than the average of the histogram is detected as a moving object. In this example, the bins enclosed in dark gray and bright gray have values larger than the average of the histogram. Correspondingly, in Fig. 6(a), the rectangle colors of the detection results are dark gray and bright gray.

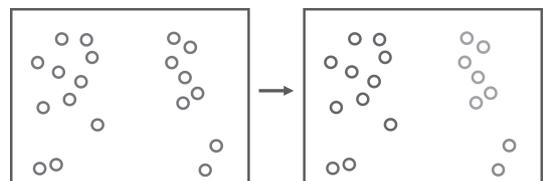
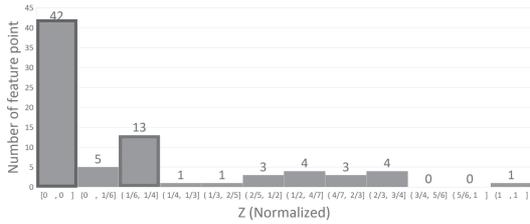


Fig. 5 Labeling feature points.



(a) An example of a scene



(b) Histogram of Z in the scene

Fig. 6 An example of creating a histogram.



Fig. 7 Stereo camera mounted on a pan-tilt head.

Table 1 Specifications of the Bumblebee2.

Model	BB2-03S2C-38
Resolution	648 × 488
Maximum frame rate	48 fps
Pixel size	7.4 μm
Focal length	3.8 mm

3. Experimental Results

To verify the validity of the proposed method, we carried out experiments.

In the first experiment, we used the Point Grey Research Bumblebee2 and KONOVA smart pan-tilt HEAD as shown in Fig. 7. Table 1 shows the specifications of the stereo camera. In the experiment, the stereo camera was panned left and right at 7°/s. Moving objects of interest were two pedestrians. They walked as shown in Fig. 8. The performance of the method was compared with [12]. The experiment was conducted off-line to compare with the same scene as in [12]. Figures 9 and 10 compare the results of the proposed method and the previous method [12] of moving object detection. The circles in these figures are feature points. Figure 9 shows that two pedestrians were successfully detected. The rectangles in these figures are

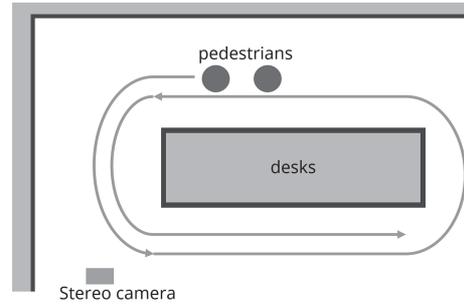


Fig. 8 Pedestrians movement.

Table 2 Results of detecting two pedestrians.

	the proposed method	[12]
T. P. rate	93.8%	75.4%
F. P. rate	6.25%	24.6%
F. N. rate	15.1%	18.9%
Precision	93.6	75.4
Recall	86.1	80.0
F value	89.8	77.6

Table 3 Detection results in scene 1.

motion		Precision	Recall	F value
camera	pedestrian	80.0	95.2	87.0
pan	horizontal	89.2	80.0	81.5
pan	forward	89.4	91.1	90.0
track	forward	82.5	84.3	83.4
dolly	forward			

Table 4 Detection results in scene 2.

motion		Precision	Recall	F value
camera	pedestrian	72.0	65.5	68.6
pan	horizontal	91.2	72.2	80.1
pan	forward	87.5	73.4	79.8
track	forward	84.8	67.6	75.2
dolly	forward			

the detection results of moving objects. In comparing Fig. 9 and Fig. 10, we found that there are many erroneous detections in Fig. 10 as compared with Fig. 9. Table 2 shows the true positive (T. P.) rate, the false positive (F. P.) rate, the false negative (F. N.) rate, precision, recall, and F value. As compared to [12], the proposed method has a lower false detection rate. We consider that the less erroneous detection of the proposed method contributed to the better performance. However, both methods failed to detect pedestrians when a) the target pedestrian was hidden due to occlusion, and b) the distance between the camera and pedestrian was too close to obtain the 3D coordinates.

As another experiment, moving object detection in the case of other movements of the camera and the moving object was performed in two locations. The images of the scenes at the two locations are shown in Fig. 11. Three kinds of camera movements: pan, track (left/right motion along the vertical direction of the optical axis), and dolly (forward/backward motion along the direction of the optical axis) were verified in the experiment.

The detection target was a pedestrian. Experimental results on movements of the camera and the moving object are shown in Tables 3 and 4. A slider was used for translational movement of the camera. The speed of the translational movement is 20 mm/s. It can be seen from the F value that good results were obtained in both scenes. The proposed method is effective in

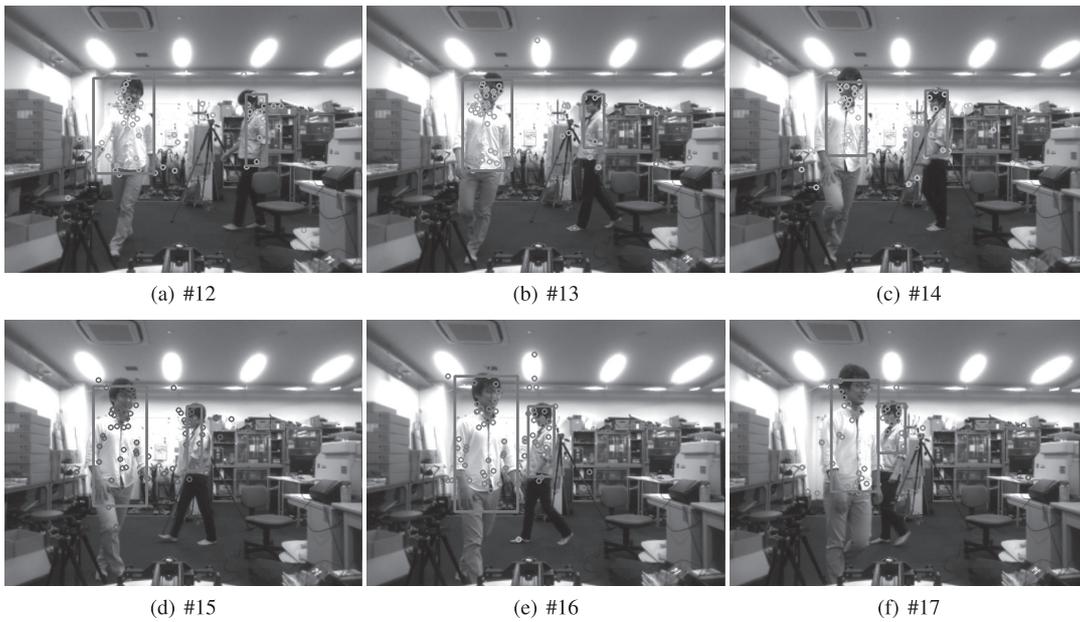


Fig. 9 Results of detection by the proposed method.

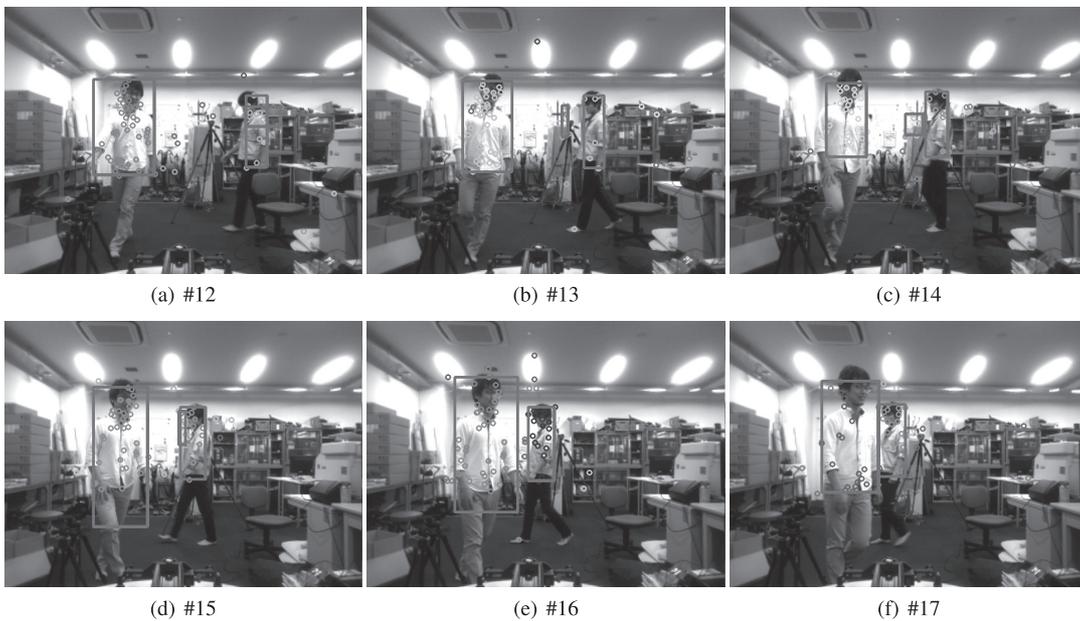


Fig. 10 Results of detection by [12].

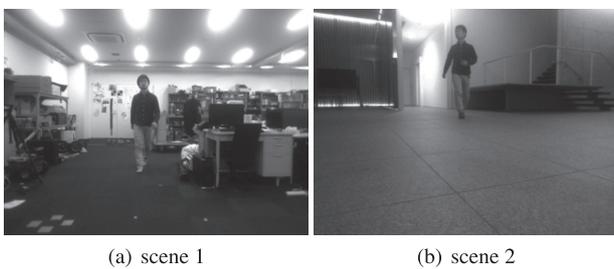


Fig. 11 The images of the scenes at two experimental locations.



Fig. 12 A trolley with cardboard boxes.

the environment like scene 2 which is darker, farther, and plain background with few feature points than in scene 1. However, the results in scene 2 are worse than that in scene 1. This is because the size of the pedestrian on the image is smaller and the moving distance between the frames is shorter, as the pedestrian

is farther from the camera.

Finally, we conducted experiment to detect moving objects other than pedestrians. The detection target was a trolley with cardboard boxes as shown in Fig. 12. The camera was moved back and forth at 20 mm/s, and the trolley was advanced towards the camera. The detection results are shown in Table 5.

Table 5 Detection results of trolley.

Precision	Recall	F value
79.8	90.9	85.0

It is found that the results are similar to the ones of the detection of a pedestrian. Therefore, this method is applicable to detection of targets other than pedestrians.

4. Conclusion

In this paper, we proposed a method for detecting moving objects using optical flow with a stereo camera. The method is not affected by the camera's motion and orientation. In the experiments, we demonstrated that the proposed method can reduce erroneous detection. We also showed that it is possible to detect moving objects in various combinations of movement of camera and moving object. Furthermore, the experiment showed that the proposed method is effective even in an environment with a plain background with few feature points. We conducted a detection experiment of moving objects other than pedestrians and showed that the proposed method can detect versatile moving objects.

In the future, it will be necessary to conduct experiments in other scenes to confirm this method's effectiveness. It is also necessary to consider the effect of the moving speed of the camera and the moving object. Furthermore, the proposed method will be evaluated through experiments using many pedestrians in an outdoor environment.

References

- [1] H. Sajid and S.-C.S. Cheung: Background subtraction for static & moving camera, *Proc. of IEEE International Conference on Image Processing*, pp. 4530–4534, 2015.
- [2] D. Zamaliev, A. Yilmaz, and J.W. Davis: Exploiting temporal geometry for moving camera background subtraction, *22nd International Conference on Pattern Recognition*, pp. 1200–1205, 2014.
- [3] Z. Hu and U. Uchimura: Tracking cycle: A new concept for simultaneous tracking of multiple moving objects in a typical traffic scene, *Proc. of IEEE Intelligent Vehicles Symposium*, pp. 233–239, 2000.
- [4] G.R. Rodriguez-Canosa, S. Thomas, J.D. Cerro, A. Barrientos, and B. MacDonald: A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVs) using a single camera, *Remote Sens.*, Vol. 4, No. 4, pp. 1090–1111, 2012.
- [5] G. Klein and D. Murray: Parallel tracking and mapping for small AR workspaces, *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234, 2007.
- [6] N. Takeda, M. Watanabe, and K. Onoguchi: Moving obstacle detection using residual error of FOE estimation, *Proc. of IROS '96*, pp. 1642–1647, 1996.
- [7] D. Yuan and Y. Yu: Moving object detection under moving camera by using normal flows, *Proc. of IEEE International Conference on Robotics and Biomimetics*, pp. 2517–2522, 2014.
- [8] D. Mochizuki, Y. Yano, T. Hashiyama, and S. Okuma: Pedestrian detection with a vehicle camera using fast template matching based on background elimination and active search, *Electronics and Communications in Japan (Part II: Electronics)*, pp. 115–126, 2007.
- [9] L. Zhao and C.E. Thorpe: Stereo- and neural network-based pedestrian detection, *IEEE Trans. on Intelligent Transportation Systems*, pp. 148–154, 2000.
- [10] J.-S. Kang, J. Kim, and M. Lee: Advanced driver assistant sys-

tem based on monocular camera, *Proc. of IEEE International Conference on Consumer Electronics*, pp. 55–56, 2014.

- [11] G. Yana, M. Yub, Y. Yub, and L. Fan: Real-time vehicle detection using histograms of oriented gradients and AdaBoost classification, *Optik – International Journal for Light and Electron Optics*, Vol. 127, pp. 7941–7951, 2016.
- [12] T. Toda, G. Masuyama, and K. Umeda: Detecting moving objects using optical flow with a moving stereo camera, *Proc. of MOBIQUITOUS 2016*, pp. 35–40, 2016.
- [13] A. Talukder and L. Matthies: Real-time detection of moving objects from moving vehicles using dense stereo and optical flow, *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3718–3725, 2004.
- [14] P.F. Alcantarilla, J. Nuevo, and A. Bartoli: Fast explicit diffusion for accelerated features in nonlinear scale spaces, *Proc. of British Machine Vision Conference 2013*, pp. 13.1–13.11, 2013.
- [15] M.A. Fischler and R.C. Bolles: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, Vol. 24, pp. 381–395, 1981.

Tetsuro TODA



He received a B.S. degree in Precision Mechanical Engineering from Chuo University, Japan, in 2016. Since 2016, he has been enrolled in the Course of Precision Engineering at Chuo University. His research interests include image processing. He is a member of the JSME and RSJ.

Gakuto MASUYAMA



He received a B.S. degree in Mechanical and Aerospace Engineering from Nagoya University, Japan, in 2005; he also earned M.S. and Ph.D. degrees in Engineering from the University of Tokyo, Japan, in 2007 and 2013, respectively. Since 2013, he has been an Assistant Professor in the Department of Precision Mechanics at Chuo University. His research interests include perceptual information processing and intelligent robotics. He is a member of the IEEE and RSJ.

Kazunori UMEDA (Member)



He received B.S., M.S., and Ph.D. degrees in Precision Mechanics from the University of Tokyo, Japan, in 1989, 1991 and 1994, respectively. From 1994 to 1998, he was a Lecturer at Chuo University. From 1998 to 2006, he was an Associate Professor in the Department of Precision Mechanics at Chuo University, and he has been a Professor in the Department of Precision Mechanics, Chuo University since 2006. From 2003 to 2004, he served as a visiting worker with the NRC of Canada. His research interests include robot vision, range image processing, and human-machine interface. He is a member of the IEEE, RSJ, JSME, JSPE and others.