Detecting Moving Objects Using Optical Flow with a Moving Stereo Camera

Tetsuro Toda
Graduate School of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
+81-3-3817-1845
toda@sensor.mech.chuo-u.ac.jp

Gakuto Masuyama
Faculty of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
+81-3-3817-1845
masuyama@mech.chuo-u.ac.jp

Kazunori Umeda
Faculty of Science and Engineering, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
+81-3-3817-1826
umeda@mech.chuo-u.ac.jp

ABSTRACT

The detection of moving objects has become an important field of research for mobile robots. It is difficult to detect moving objects from a moving camera because moving objects and the background can appear to move. This paper proposes 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 has been evaluated through experiments using two pedestrians in an indoor environment.

CCS Concepts

• **Computing methodologies→Object detection** • **Computing methodologies→Vision for robotics**

Keywords

Stereo camera; Optical flow

1. INTRODUCTION

To achieve the automatic operation of autonomous and automobile robots, research regarding the detection of moving objects has become increasingly important. This paper presents a system of detecting moving objects by using images obtained from a camera mounted on objects such as cars and mobile robots.

In images obtained from a moving camera, both moving objects and the background could appear to move as shown in Figure 1. Therefore, it is not plausible to apply a simple background subtraction method to detect moving objects in the scene. On the other hand, Hu *et al*. [1] proposed a method for detecting moving objects by adaptive background subtraction, which adjusts the motion of background based on analysis of the camera's motion. Rodriguez *et al*. [2] used a Parallel Tracking and Mapping (PTAM) [3] algorithm to estimate the camera's motion parameters. Moving objects can be detected by comparing the optical flow obtained

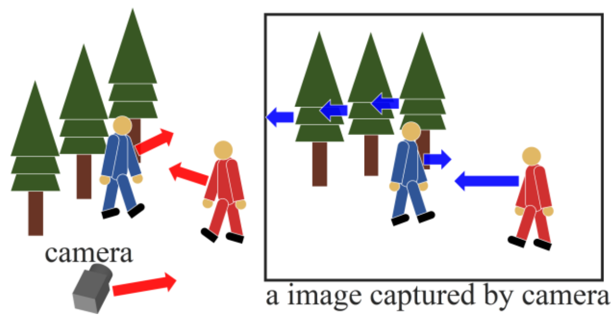


Figure 1: An illustrative image captured from a moving camera. Red arrows represent movements of the camera and pedestrians; Blue arrows represent movements of objects in the image.

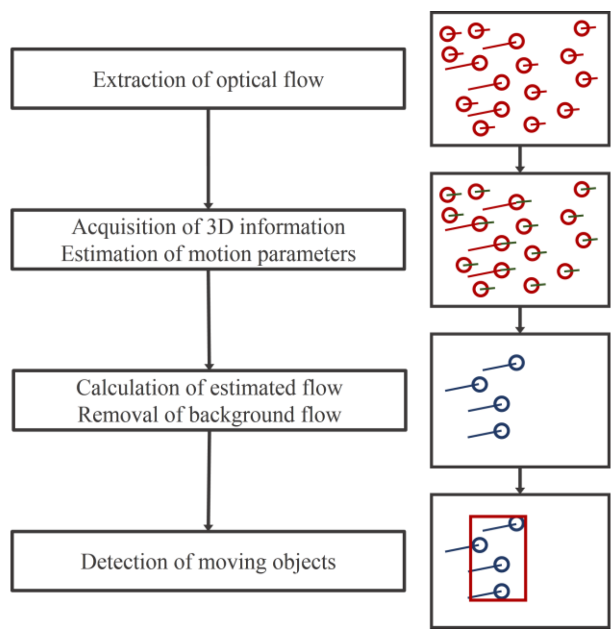


Figure 2: Flow of the proposal method

MOBIQUITOUS '16 Adjunct Proceedings, November 28-December 01, 2016, Hiroshima, Japan

DOI: http://dx.doi.org/10.1145/3004010.3004016

from the images and an artificial optical flow obtained motion parameters. These two methods are designed for cameras mounted on cars and an Unmanned Aerial Vehicle (UAV) that have limited camera motion and orientation. Takeda *et al*. [4] proposed a method using Focus of Expansion (FOE) of the optical flow. This approach is not limited to the camera's movement and orientation. However, it cannot detect moving object when the camera and the moving object are moving in the same direction. As a method of a different approach, Mochizuki *et al*. [5] propose the method using template matching based on background elimination. Another method based on stereo camera and Neural Network was proposed by Zhao *et al*. [6]. However, these methods using template matching or Neural Network is limited in a target because there is a need to prepare the target of template.

We propose a method that takes a different approach—using a stereo camera to detect moving object based upon three-dimensional coordinates that are not limited to the camera's motion and orientation. The flow of the system is described in Figure 2. 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, which includes falsely matched flows. Our purpose is to remove all optical flows except for moving object flow. Second, camera motion parameters are estimated using three-dimensional information of each feature point. Finally, the optical flow is estimated when it is assumed that there is no moving object in the images. Hereinafter, the estimated optical flow denotes the estimated flow. The background flow is removed by comparing the estimated flow and the optical flow obtained from the image. The moving object is then 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 our method. In Section 3, we describe moving object-detection experiments in to show the validity of our research. Finally, our conclusions and goals for future work are shown in Section 4.

2. DETECTING MOVING OBJECTS

2.1 Extracting Optical Flow

The extraction of optical flow is shown in Figure 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 $\boldsymbol{F}_k$ is a movement vector from p_k^{t-1} to p_k^t, where p_k^{t-1} and p_k^t are feature points in f^{t-1} and f^t, respectively, and $k = 1, \cdots, n$ (n: number of matched feature points). The feature points are extracted with Accelerated KAZE (AKAZE) [7]. 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.

2.2 Estimating Camera Motion Parameters

Figure 4 shows the camera motion parameters. The translation vector $\boldsymbol{t}$ and the rotation matrix R are represented as

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

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

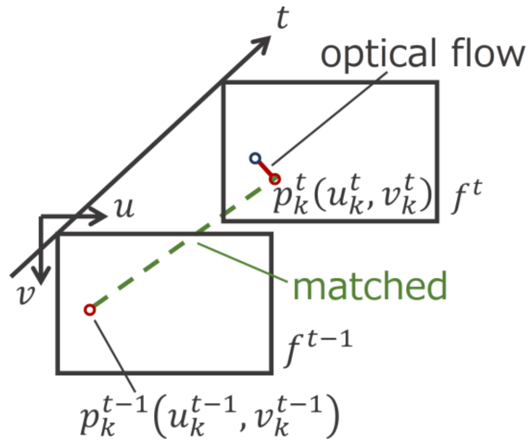


Figure 3: Extraction of optical flow

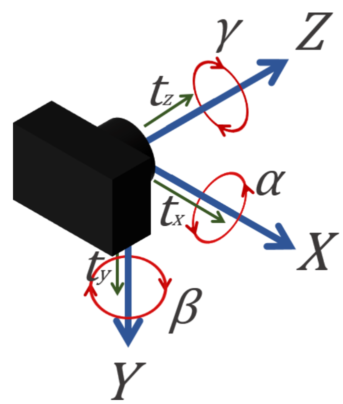


Figure 4: Camera motion parameters

where R is a linearized rotation matrix approximated by the rotation angle of the camera that is assumed to be minute; and α, β, and γ are the rotation angles around the X, Y, and Z axes, as shown in Figure 4.

The three-dimensional coordinates of feature points p_k^t and p_k^{t-1} are $\boldsymbol{P}_k^t = [X_k^t \quad Y_k^t \quad Z_k^t]^T$ and $\boldsymbol{P}_k^{t-1} = [X_k^{t-1} \quad Y_k^{t-1} \quad Z_k^{t-1}]^T$, respectively. Then, the relationship of $\boldsymbol{P}_k^t$ and $\boldsymbol{P}_k^{t-1}$ is described as follows:

$$\boldsymbol{P}_k^{t-1} = R\boldsymbol{P}_k^t + \boldsymbol{t}\,. \tag{3}$$

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

$$\boldsymbol{P}_k^{t-1} - \boldsymbol{P}_k^t = R'\boldsymbol{P}_k^t + \boldsymbol{t}\,, \tag{4}$$

where R' has off-diagonal elements identical to those of R, and its diagonal elements are 0. Then, the relationship between $\boldsymbol{P}_k^t$, $\boldsymbol{P}_k^{t-1}$ and the parameters is described as

$$A_k\boldsymbol{\xi} = \boldsymbol{b}_k, \tag{5}$$

Where

$$A_k = \begin{bmatrix} 0 & Z_k^t & -Y_k^t & \\ -Z_k^t & 0 & X_k^t & I_3 \\ Y_k^t & -X_k^t & 0 & \end{bmatrix}, \tag{6}$$

$$\boldsymbol{\xi} = [\alpha \quad \beta \quad \gamma \quad t_x \quad t_y \quad t_z]^T\,, \tag{7}$$

$$\boldsymbol{b}_k = \boldsymbol{P}_k^{t-1} - \boldsymbol{P}_k^t\,. \tag{8}$$

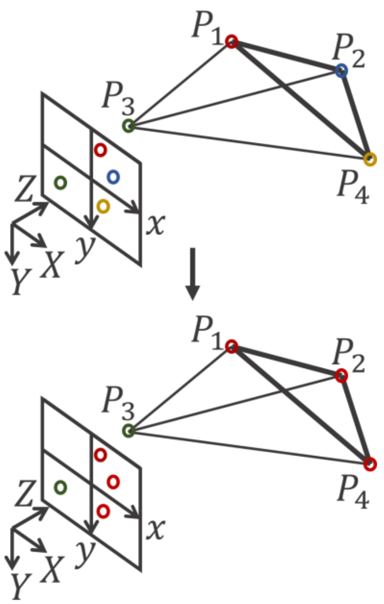
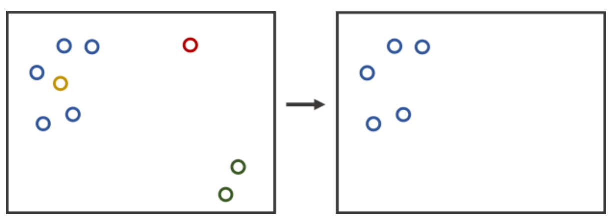
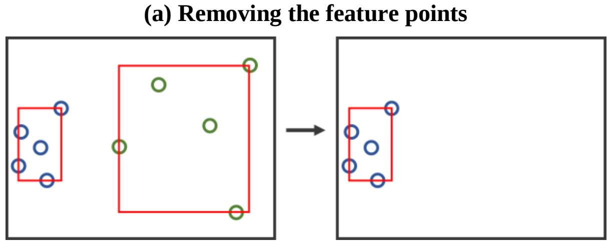


Figure 5: Camera motion parameters



(a) Removing the feature points



(b) Removing detection rectangle

Figure 6: Removing false detections

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

$$\boldsymbol{\xi} = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}^+ \begin{bmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_n \end{bmatrix}, \tag{9}$$

where

$$\begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}^+ = \left(\begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}^T \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix} \right)^{-1} \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}. \tag{10}$$

However, the matching results have outliers due to false matching etc.; thus, outliers are removed by Random Sample Consensus (RANSAC) [8].

2.3 Calculating estimated flow

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

$$x_k^t = \frac{X_k^t}{Z_k^t}, \tag{11}$$

$$y_k^t = \frac{Y_k^t}{Z_k^t}. \tag{12}$$

The estimated flow $\boldsymbol{F'}_k = [\dot{x}_k \quad \dot{y}_k]^T$ is calculated as shown below:

$$\boldsymbol{F'}_k = \begin{bmatrix} \dot{x}_k \\ \dot{y}_k \end{bmatrix} = \begin{bmatrix} \frac{\partial x_k}{\partial t} \\ \frac{\partial y_k}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{X_k^{t-1}}{Z_k^t} - \frac{X_k^t Z_k^{t-1}}{{Z_k^t}^2} \\ \frac{Y_k^{t-1}}{Z_k^t} - \frac{Y_k^t Z_k^{t-1}}{{Z_k^t}^2} \end{bmatrix}, \tag{13}$$

where X_k^{t-1}, Y_k^{t-1}, and Z_k^{t-1} are represented by equation (3) as follows:

$$X_k^{t-1} = X_k^t - \gamma Y_k^t + \beta Z_k^t + t_x, \tag{14}$$

$$Y_k^{t-1} = \gamma X_k^t + Y_k^t - \alpha Z_k^t + t_y, \tag{15}$$

$$Z_k^{t-1} = -\beta X_k^t + \alpha Y_k^t + Z_k^t + t_z. \tag{16}$$

Therefore, the estimated flow $\boldsymbol{F'}_k$ is described as

$$\boldsymbol{F'}_k = \begin{bmatrix} -\gamma y_k^t + \beta + \frac{t_x}{Z_k^t} - x_k^t \left(-\beta x_k^t + \alpha y_k^t + \frac{t_z}{Z_k^t} \right) \\ -\gamma x_k^t + \alpha + \frac{t_y}{Z_k^t} - y_k^t \left(-\beta x_k^t + \alpha y_k^t + \frac{t_z}{Z_k^t} \right) \end{bmatrix}. \tag{17}$$

Then, a point $p'^{t-1}_k = [u'^{t-1}_k \quad v'^{t-1}_k]^T$ is estimated by equation (17).

2.4 Removing background flow

The Euclidean distance between p_k^t and p'^{t-1}_k is represented as d_k. When d_k is lower than threshold d_{th}, optical flow $\boldsymbol{F}_k$ is removed as background flow.

2.5 Detecting moving objects

In order to detect moving objects individually, feature points $\{p^t\} = \{p_0^t, \dots, p_n^t | d_k > d_{th}\}$ are labeled as shown in Figure 5. When the Euclidean three-dimensional distance D_{ij} between an arbitrary feature point p_i^t and p_j^t is lower than threshold D_{th}, p_i^t and p_j^t are given the same label. The set of feature points with the same label l is expressed as $\{p_l^t\}$.

The detection results are represented as a small rectangle that surrounds the feature points of the same label. Therefore, in order to reduce false detections, this process is performed for each label as follows:

- If the number of feature points with the same label is small, the detection is removed (see Figure 6(a));
- It is assumed that the lager the detection area rectangle becomes, the more feature points with the same label there will be. Therefore, the detection is removed when the number of feature points with the same label is small with respect to the area (see Figure 6(b)).

The detection is removed as a false detection if it does not satisfy the following condition:

$$|p_l^t| \geq \frac{S_l}{S_{th}} + m_{th}, \tag{18}$$

where $|p_l^t|$ is the number of feature points given label l, S_l is the area of the detection rectangle of label l, S_{th} is threshold with respect to S, and m_{th} is minimum required number of feature points not removed as false detections.

3. EXPERIMENTAL RESULTS

To verify the validity of our proposed method, we carried out an experiment. Moving objects of interest were two pedestrians, and a stereo camera panned left and right.

In this experiment, we used Point Grey Research Bumblebee2 camera and KONOVA smart pan-tilt HEAD as shown in Figure 7. Table 1 shows the specification of the stereo camera, and Table 2 shows the specification of the pan-tilt head. The pan-tilt speed was 7 °/s in the experiment. Thresholds were set as shown in Table 3.

Figure 8 shows the result of moving object detection using our proposed method. Figure 8(a) shows input from the current image. Figure 8(b) shows the result of extracting the optical flow. The red circles are the feature points $\{p_0^t, \ldots, p_n^t\}$, the red lines are the optical flows and green lines are the estimated flows. Figure 8(c) shows the results of detecting moving objects. Circles of the same color are feature points with the same label, and red rectangles are the detection area. It can be seen that the two pedestrians were successfully detected.

Table 4 shows the true positive rate, the false positive rate, and the false negative rate. The false negative rate was somewhat high; however, the distance value of moving objects could not be acquired in many sequences, e.g., when occlusion occurred (Figure 9(a)) or when a moving object was too close to the stereo camera (Figure 9(b)). As seen in Table 4, the false positive rate was high. This is because false positives were formed by optical flows occurring around the moving objects shown in Figure 10.

Table 1: Specification of Bumblebee2

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

Table 2: Specification of smart pan-tilt HEAD

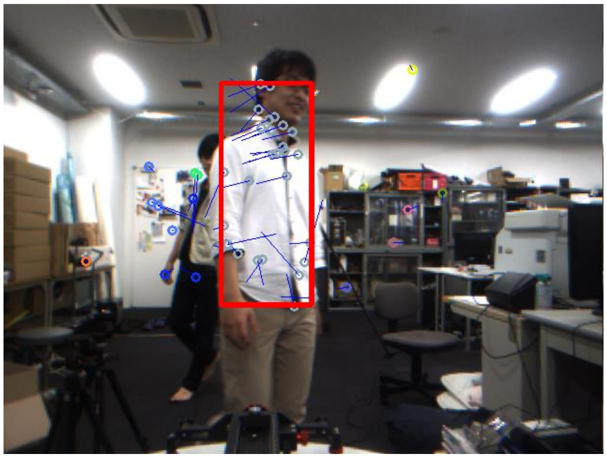
Minimum pan speed	3 °/s
Maximum pan speed	7 °/s

Table 3: Configuration of thresholds

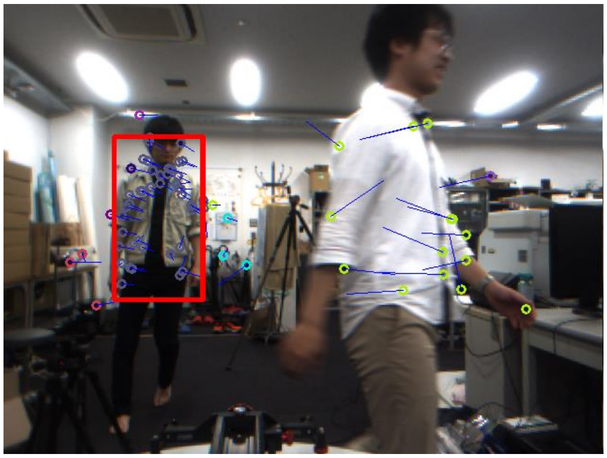
Threshold	d_{th}	D_{th}	S_{th}	m_{th}
Value	4 pixels	0.8 m	3600 pixels^2	3

Table 4: detection results

Frames	True positive rate	False positive rate	False negative rate
37	81.1%	26.4%	19.0%



(a) A scene of occulusion

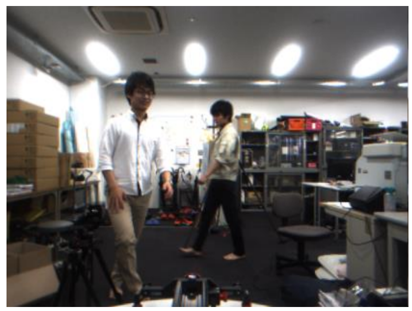


(b) A scene too close to the camera

Figure 9: False negative detection

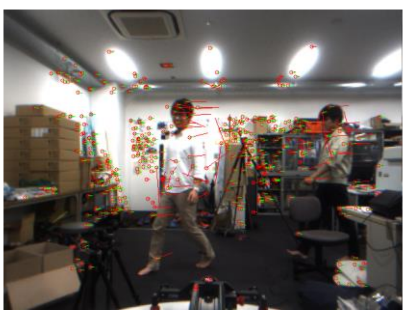
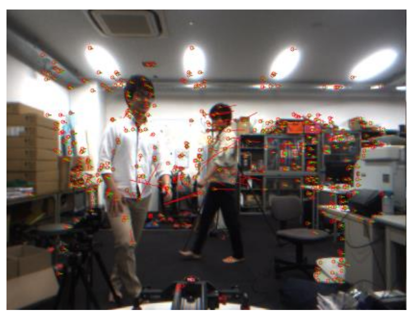
Figure 10: False positive detection

Figure 7: Stereo camera mounted on pan-tilt head

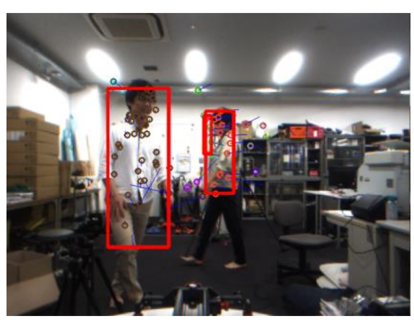
$t = 3$ $t = 9$ $t = 11$

(a): Input current image

$t = 3$ $t = 9$ $t = 11$

(b): Result of extraction of optical flow

$t = 3$ $t = 9$ $t = 11$

(c): Result of detection

Figure 8: Result of experiment

4. CONCLUSION

In this paper, we proposed a method for detecting moving objects using optical flow with a stereo camera. It was not limited to the camera motion and orientation. In the experiment, our method successfully detected two pedestrians.

In the future, we will work to reduce false detections. Furthermore, the proposed method will be evaluated through experiments using many pedestrians in an outdoor environment.

5. REFERENCES

[1] Hu, Z. and Uchimura, U. 2000. Tracking cycle: a new concept for simultaneous tracking of multiple moving objects in a typical traffic scene. In *Intelligent Vehicles symposium*, 233-239.

[2] Rodriguez-Canosa, G R, Thomas, S, Cerro, J D, Barrientos, A, and MacDonald, B. 2012. A real-time method to detect and track moving objects(DATMO) from unmanned aerial vehicles (UAVs) using a single camera. In *Remote Sensing*, 1090-1111. DOI= http://dx.doi.org/10.3390/rs4041090.

[3] Klein, G and Murray, D. 2007. Parallel Tracking and Mapping for Small AR Workspaces. In *the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 225-234. DOI= http://dx.doi.org/10.1109/ISMAR.2007.4538852.

[4] Takeda, N, Watanabe, N, and Onoguchi, K. 1996. Moving obstacle detection using residual error of FOE estimation. In *IROS96*, 1642-1647. DOI= http://dx.doi.org/10.1109/IROS.1996.569032.

[5] Mochizuki, D, Yano, Y, Hashiyama, T, and Okuma, S. 2007. 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). 115-126. DOI= http://dx.doi.org/10.1002/ecjb.20393.

[6] Zhao, L and Thorpe, C E. 2000. Stereo- and neural network-based pedestrian detection. In *IEEE Transactions on Intelligent Transportation Systems*. 148-154. DOI= http://dx.doi.org/10.1109/6979.892151.

[7] Alcantarilla, P F, Nuevo, J, and Bartoli, A. 2013. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *BMVC13*. 13.1-13.11. DOI= http://dx.doi.org/10.5244/C.27.13.

[8] Fischler, M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 381-395. DOI= http://dx.doi.org/10.1145/358669.358692.