Visual Odometry from Floor Images by using Accelerated KAZE Features

Tomohiro Morio^a, Shingo Aramaki^a, Shota Nakashima^{a,*}

^aYamaguchi University

2-16-1, Tokiwadai, Ube, Yamaguchi, 755-8611, Japan

*Corresponding Author: s-naka@yamaguchi-u.ac.jp

Abstract

For the estimation of self-position in an autonomous robot, there is a requirement to measure the two-dimensional movement. Therefore, a self-position estimation method using the floor image photographed with a CCD camera is proposed. In this paper, the self-position is estimated by detected keypoints from the image photographed before and after movement. AKAZE is used for the extraction of the features. Using the extracted existing frame by the past movement prediction, the movement position of 1 frame ahead of each keypoints can be predicted. Therefore, using the movement prediction of the feature point, in comparison with feature matching method, a method to speed up processing is proposed.

Keywords: Visual Odometry, Odometry measurement, AKAZE features, Descriptor Matching.

1. Introduction

For the estimation of self-position in an autonomous robot, there is a requirement to measure the two-dimensional movement. The measurement method includes the use of collected odometry wheel data by using a rotary encoder. From the uses of odometry wheel data, the movement can be determined based on the quantity of rotation of the right and left wheel. However, it is thought that an error occurs to the movement because of the changes of the wheel diameter as dirt attaches to the wheel due to the influence of the wheel sliding⁽¹⁾ and contact with the floor. Hence, without using the odometry wheel data, a self-position estimation method using the captured image is proposed. Self-position estimation method using captured image of an autonomous robot includes the uses of template matching method and the

uses of keypoints. In the present, some examples of template matching method uses in self-position estimation is Nagai's Downward Facing Camera method⁽²⁾, Savan's Ground Vehicle method⁽³⁾ and Andrew's Real-Time Stereo Visual Odometry method⁽⁴⁾. On the other hand, the method using template matching have a need in controlling the illumination of light. When the changes of color occur because of the disturbance of illumination, image matching cannot be performed on the floor image⁽⁵⁾. Thus, a self-position estimation method by using an effective keypoints method on the color changes is proposed. Applied examples of the self-position estimation using the keypoints method include research from Mark⁽⁶⁾, Chen⁽⁷⁾, Raul^(8,9), Stephen⁽¹⁰⁾, Brian⁽¹¹⁾, Kurt⁽¹²⁾, Timothy⁽¹³⁾, Christian⁽¹⁴⁾, Mikael⁽¹⁵⁾, and Julian⁽¹⁶⁾. In order to detect the keypoints, studies from Mark uses the chessboard corner, then uses Lucas-Kanade method for the determination of keypoints⁽⁶⁾. However, it is necessary to place a chessboard in the floor and cannot be applied for real environment uses. The studies from Chen, Raul, Stephen, Brian, Kurt, Timothy, Christian, Mikael, and Julian photograph landmarks such as a building or a structure, then extracted the keypoints, corners or edges of the landmark's area using descriptor such as SIFT, SURF, FAST or ORB. The feature of an image is extracted from the image using descriptor⁽⁷⁻¹⁶⁾. Furthermore, this method corresponds and tracks the keypoints by comparing the details on both before and after movement of the image. However, since the method uses keypoints from landmarks, there is a possibility that the landmark keypoints will lose sight due to the environment lighting or disturbance. In addition, the calculation cost for the feature data is required⁽²⁾.

Therefore, a highly competent in changes of brightness and rotation of AKAZE's feature is used. Regarding this, a method using only the keypoints of the floor image and the

prediction of the keypoints' coordinate of the afterward frame from the movement result of the previous frame is proposed. Moreover, from the joining result of the keypoints extracted, this method can reduce the calculation cost.

2. Proposed System

2.1 System Summary

For the measurement of the movement, a monocular camera pointed perpendicularly to the floor is use. As shown in Fig. 1, the image is photographed as the camera moves parallel to the floor. When moving, the self-position estimation is performed and constructed from the acquired image photographed continuously. Next, from the keypoints obtained from two consecutive photographed frames and from the image, the feature keypoints is extracted. The keypoints of the two images is matched and the correspondence point between the images is determined. As shown in Fig. 1, as the camera faces perpendicularly to the floor, when the height between the camera and the floor is d , the half angle of the camera is θ and the number of pixel in the x direction of the image sensor is w , and when the considerate point (x^f, y^f) moves to (x^{f+1}, y^{f+1}) , the movement of x direction and y direction can be express as respectively and shown in the following expression.

$$\begin{bmatrix} v_x^{f+1} \\ v_y^{f+1} \end{bmatrix} = \frac{2d \tan \theta}{w} \begin{bmatrix} x^{f+1} - x^f \\ y^{f+1} - y^f \end{bmatrix} \quad (1)$$

2.2 Prediction of Next Frame

For the self-position estimation using the keypoints, when using feature quantities obtained from matching method is performed, calculation cost is necessary. In conjunction, without using the feature from the keypoints, using the movement calculated from the current frame and one frame before it, the next position of the keypoints is

Fig. 1. Motion measurement using floor image.

predicted. A method reducing the calculation cost corresponds to the keypoints obtained from the photographed image of the after movement is also proposed. This method uses the movement of the obtained current frame and the movement of the one frame before it to estimate the next frame. When there is inertia on the camera capturing the floor image as the camera move gently parallel to the floor, and when the frame rate is assume sufficiently fast, the gap between the movement obtained from the current frame and the next frame can be considered as the gap between previous one frame and the current frame. Therefore, as the camera moves on the x-axis and y-axis and when the one frame before, the current frame, and the next frame after is the obtained considerate point, the coordinate of the existing frame keypoints can be define as (x^f, y^f) , the movement in x-axis and y-axis between current frame and one frame before can be define as v_x^f and v_y^f respectively, the angle of rotation when the rotary motion is perform can be define as θ^f , the movement position in x-axis and y-axis of the keypoints in the next frame can be define as x'^{f+1} , y'^{f+1} correspondingly and can be expressed in the following expression.

$$\begin{bmatrix} x'^{f+1} \\ y'^{f+1} \end{bmatrix} = \begin{bmatrix} x^f \\ y^f \end{bmatrix} - \begin{bmatrix} \cos \theta^f & \sin \theta^f \\ -\sin \theta^f & \cos \theta^f \end{bmatrix} \begin{bmatrix} v_x^f \\ v_y^f \end{bmatrix} \quad (2)$$

In this experiment, as the movement of x-axis in a forward direction is considered, the prediction can be done by the following expression.

$$\begin{bmatrix} x'^{f+1} \\ y'^{f+1} \end{bmatrix} = \begin{bmatrix} x^f - v_x^f \\ y^f - v_y^f \end{bmatrix} \quad (3)$$

After the prediction, the distance of the prediction point and the keypoints of the ground truth is measured. The correspondence predicted keypoints of the current frame is then associated with the nearest keypoints. On this time, the predictive point outside the next frame is excluded. Thus, it is possible that the estimated predictive point result of the next frame movement position is correspondence to the original coordinate and also possible for the unnecessary estimated point corresponds to the original coordinate. From the above explanation, without using its feature, the extracted keypoints of the current frame and the next frame can be associated.

2.3 Remove Incorrectly Related Points

Current keypoints can relate to next flame keypoints without compute their features shown by section 2.2. However, the proposed method sometimes relate to incorrectly keypoints because of some current frame keypoints may vanish in next flame if lighting changed. These incorrectly related points are effected to measurement results. Our proposed method use standard deviation for remove word related points. Firstly, calculate all x-axis and y-axis movement using all related points form current and next flame. Secondly, calculate mean and standard deviation from all x-axis and y-axis movement. Thirdly, define average of x or y axis movement as a_m , define standard deviation of x or y axis as s_m and define measured movement using one related points as m , proposed method can detect correct related point from following threshold of $a_m - s_m \leq m \leq a_m + s_m$. Finally, define average of x or y axis movement as a_n , define standard deviation of x or y axis as s_n and define measured movement using one related points as n , proposed method can detect correct related point again from following threshold of $a_n - s_n \leq n \leq a_n + s_n$. Doing so, measurement result can enhance precision.

3. Evaluation Experiments

3.1 Experimental Environment and Method

For the experiment setup, STC-SB32POEHS (produced by OMRON SENTECH) was used for the CCD camera which capture the floor image, moreover MV-0614N (produced by NS-Lighting) was used for the lens which attached to CCD camera. The camera was installed facing downward 100 mm vertically with floor. The proposed system used light shielding plate for box and box's width is 296 mm and measure 236 mm in depth. The box is installed measure 3 mm in gap from floor. In addition, proposed system used linear guide for move system linearly. For the

Fig. 3. Captured floor image.

Fig. 4. Detected keypoints from image.

lighting, two OPB-20015W2 (produced by OPTEX FA) was used. The direction of radiation of the LEDs are parallel to the floor and towards to the center of image which can illuminate asperity and scratch of the floor. In addition, two LEDs illuminate are installed on both side of the box. Figure 2 shows experimental environment of our proposed system. For a subject, P-72 (produced by TAJIMA ROOFING) of the floor tile was used. Figure 3 shows captured floor image using proposed system and Fig. 4 shows detected keypoints from captured floor image using AKAZE descriptor. In addition, for the PC, HP Compaq Elite 8300 CMT (produced by Hewlett-Packard, Intel Core i7 3770 used) was used for compute and estimate move. MLS-50-1080C-2000 (produced by MICROTECH LABORATORY) was used for the wire encoder which measured ground truth. This experiment took pieces of floor image which we photographed every 1 mm for the x-axis direction by using wire encoder. Images was chosen and tested in 3 kinds of movement models, equal in speed and includes sigmoid function which accelerates and decelerates. Below shows the 3 kinds of movement models which was experimented.

I) 200 mm by equal speed model, 4 mm interval per 1 flame

II) 10 mm by gain 1.25, sigmoid function model

III) 10 mm by gain 1.75, sigmoid function model

Firstly, the proposed method detected keypoints from before and after the pictures move using movement models.

Fig. 2. Experimental environment.

Fig. 5. Computed system movement including incorrectly related points.

Fig. 6. Removed incorrectly related points.

Secondly, computed features from pictures and keypoints only of first time to get measurement result. Thirdly, features and compute movement is matched. Fourthly, predicted next flame movement using matching result. Finally, keypoints from after and next flame is detected and matched each keypoints using predicted result. Figure 5 shows captured floor image including line which described by computed system movement between before and after moved picture. In Fig. 5, green lines show correctly related points and red lines show incorrectly related points. In addition, Fig. 6 shows the result of removing the selected lines from Fig. 5 which described as incorrectly related points by using standard deviation.

Equal speed model estimated movement 150 times, and sigmoid speed model estimated model 190 times. The mean of the movement every flame was calculated. In addition, the mean estimate time every movement models was measured.

3.2 Experimental Results

Relations of the movement with a flame of 3 models each is shown in Fig. 7(a) to (c). In addition, the standard division with the processing time of 3 experimented models is shown in Fig. 8(a) to (c). According to Fig. 7(a) and (b), both results of measurement move are follow-up to ground truth. However, according to Fig. 7(c), only AKAZE features method could follow-up to ground truth, but AKAZE features and estimate method cannot follow-up to ground

(a) Equal speed model.

(b) Sigmoid model, gain 1.25.

(c) Sigmoid model, gain 1.75.

Fig. 7. Result of estimate measurement.

truth. This problem is thought to be causally related to previous flame's result. Proposed method estimated to move 0 mm in next flame. Our proposed method related incorrectly keypoints because of some keypoints around estimated point. To solve this problem, if the estimated point have AKAZE features, check features around the estimate point will promise a precise estimate result. According to Fig. 8(a) to (c), each model of processed time are faster than conventional method. The experimental results show if odometry robot move smoothly, our proposed system possible to estimate movement follow-up to ground truth.

4. Conclusions

In this paper, visual odometry method using AKAZE features and estimate method was proposed. To estimate odometry, the measurement of 2-D motion is necessary. Measurement of the 2-D motion has contact and non-contact system. Common uses of contact system is using wheel encoder. However, most of wheel encoder changes the wheel radial by dirt during measurement. This will cause problems and errors in the measurement. To overcome this problem, non-contact method was introduced. In this paper, method using floor images, AKAZE features and estimate method is introduced. Estimate method forecasts next flame coordinate which current flame keypoints will move from current flame and previous one flame. Using estimate keypoints, current flame's keypoints can relate to next flame's keypoints. In addition, the proposed method use standard deviation two times for compute measurement results to enhance precision. In experiment, we measured equal speed model and sigmoid model which can measure accelerate and decelerate by using proposed method. From the result of experiment, the proposed system can follow-up to ground truth if odometry robot move smoothly.

Therefore, as a future task, the uses of more measurement data is necessary to estimate next flame coordinate to enhance accuracy.

References

- (1) Michael Dille, Ben Grocholsky, and Sanjiv Singh : "Outdoor Downward-Facing Optical Flow Odometry with Commodity Sensors", Field and Service Robotics, pp. 183-193, 2010
- (2) Isaku Nagai and Keigo Watanabe : "Path tracking by a mobile robot equipped with only a downward facing camera", IEEE International Conference on Intelligent Robots and Syetems, pp. 6053-6058, 2015
- (3) Savan Chhaniyara, Pished Bunnun, Lakmal D. Seneviratne, and Kaspar Althoefer : "Optical Flow Algorithm for Velocity Estimation of Ground Vehicles: A Feasibility Study", International Journal on Smart Sensing and Intelligent Systems, Vol. 1, No. 1, pp. 246-268, 2008
- (4) Andrew Howard : "Real-Time stereo visual odometry for autonomous ground vehicles", IEEE International Conference on Intelligent Robots and Systems, pp. 3946-3952, 2008
- (5) David G. Lowe : "Object recognition from local scale-

(a) Equal speed model.

(b) Sigmoid model, gain 1.25.

(c) Sigmoid model, gain 1.75.

Fig. 8. Measurement time using estimate method.

- invariant features", The Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 1150-1157, 1999
- (6) Marc Killpack, Travis Deyle, and Charles C. Kemp : "Visual odometry and control for an omnidirectional mobile robot with a downward-facing camera", IEEE International Conference on Intelligent Robots and Systems, pp. 139-146, 2010
- (7) Kuan-Chieh Chen and Wen-Hsiang Tsai : "Vision-Based Autonomous Vehicle Guidance for Indoor Security Patrolling by a SIFT-Based Vehicle-

- Localization technique”, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, Vol. 59, No. 7, pp. 3261-3271, 2010
- (8) Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós : “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, IEEE Transactions on Robotics, Vol. 31, No. 5, pp. 1147-1163, 2015
 - (9) Raúl Mur-Artal, and Juan D. Tardós : “ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras”, IEEE Transactions on Robotics, Vol. 33, No. 5, pp. 1255-1262, 2017
 - (10) Stephen Se, David Lowe, and Jim Little : “Vision-based mobile robot localization and mapping using scale-invariant features”, IEEE International Conference on Robotics and Automation, pp. 2051-2058, 2001
 - (11) Brian Clipp, Jongwoo Lim, and Jan-Michael Frahm : “Parallel, real-time visual SLAM”, IEEE International Conference on Intelligent Robots and Systems, pp. 3961-3968, 2010
 - (12) Kurt Konolige, Motilal Agrawal, and Joan Solà : “Large Scale Visual Odometry for Rough Terrain”, Robotics Research, pp. 201-212, 2011
 - (13) Timothy D. Barfoot : “Online Visual Motion Estimation using FastSLAM with SIFT Features”, IEEE International Conference on Intelligent Robots and Systems, pp. 579-585, 2005
 - (14) Christian Forster, Zichao Zhang, Michael Gassner, Manuel Werlberger, and Davide Scaramuzza : “SVO: Semidirect Visual Odometry for Monocular and Multicamera Systems”, IEEE Transactions on Robotics, Vol. 33, No. 2, pp. 249-265, 2017
 - (15) Mikael Persson, Tommaso Piccini, Michael Felsberg and Rudolf Mester : “Robust stereo visual odometry from monocular techniques”, IEEE Intelligent Vehicles Symposium, pp. 686-691, 2015
 - (16) Julian Straub, Sebastian Hilsenbeck, Georg Schroth, Robert Huitl, A Moller, and Eckehard Steinbach : “Fast relocalization for visual odometry using binary features”, IEEE International Conference on Image Processing, pp. 2548-2552, 2013