IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 8, 2019, accepted January 19, 2019, date of publication February 22, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901008

AKAZE-Based Visual Odometry From Floor Images Supported by Acceleration Models

SHOTA NAKASHIMA[1], (Member, IEEE), TOMOHIRO MORIO[1], AND SHENGLIN MU[2], (Member, IEEE)

[1]Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube 755-8611, Japan
[2]Graduate School of Science and Engineering, Ehime University, Ehime 790-8577, Japan

Corresponding author: Shota Nakashima (s-naka@yamaguchi-u.ac.jp)

ABSTRACT To realize the self-localization of autonomous robots, methods for the 2D motion measurement of robots are required. In this research, a self-localization system using a CCD camera is proposed. In the proposed system, the self-localization is estimated by movement tracking using some keypoints detected from the floor images captured by the CCD camera. For the illumination of floor image, two LED illuminate are used. These lighting systems are installed in such a way that lit from both sides of the floor and parallel to the floor so that some minimal bumps or original veins are captured from the floor. An accelerated KAZE feature is applied in the proposed system for keypoint detection and for computing resulting in the generation of the descriptor. In the estimation, the location information from the previous step and the estimated position one frame ahead are used. Using the estimated result and computed descriptor, the proposed system matches the nearest keypoint between the estimated keypoint detected from the previous floor image and the keypoint detected from the next floor image at the beginning. The Hamming distance is employed in the proposed system to evaluate the matching. If the Hamming distance is longer than a threshold, the proposed system tries to match from the 2nd nearest keypoint. Based on an experiment in which the measurement distance and computation time are investigated, the effectiveness of the proposed method is confirmed.

INDEX TERMS Visual odometry, vision measurement, image motion analysis, odometry measurement, AKAZE features, image matching.

I. INTRODUCTION

For the estimation of self-localization 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 odometry wheel data, the movement can be determined based on the quantity of rotation of the right and left wheels. However, it is thought that an error occurs in the movement because of the changes in wheel diameters that arises 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-localization method using the captured images is proposed. The self-localization method using captured images of an autonomous robot includes the use of the template matching method and the use of keypoints. Currently, some examples of template matching methods applied in self-localization are Nagai's downward facing camera method [2], Savan's ground vehicle method [3] and Andrew's real-time stereo visual odometry method [4]. However, the method using template matching needs to control the illumination of light. When color changes occur because of the disturbance of illumination, image matching cannot be applied to the floor image [5]. For image recognition, the methods using neural network were proposed by Sakai [6], [7] and Nakayama [8]. However, because of the learning process beforehand, neural network is not suitable to be applied in visual odometry. The self-localization method by using keypoints method on the color changes is proposed. Some studies on the keypoint-based method have been conducted by Marc [9], Chen [10], Raul [11], [12], Stephen [13], Brian [14], Kurt [15], Timothy [16], Christian [17], Mikael [18], and Julian [19]. To detect the keypoints, studies from Marc use the chessboard corner, after which the Lucas-Kanade method is employed to determine keypoints [9]. However, as a

The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu.

chessboard is required to be placed on the floor, the method is difficult to apply in the field. In the studies of Chen, Raul, Stephen, Brian, Kurt, Timothy, Christian, Mikael, and Julian, landmarks such as a building or structure are photographed; then, keypoints are extracted from the corners or edges of landmarks' area using descriptors such as SIFT, SURF, FAST or ORB. The features of an image are extracted from the image using a descriptor [10]–[19]. However, the descriptor-matching method may lose the keypoints detected from the landmarks if sight of the landmark is lost, which could be affected by illumination. Therefore, based on our previous study [20], a self-localization method using feature points extracted from floor images by accelerated KAZE features (AKAZE) is proposed for movement prediction. As a method of feature point matching, the position after the movement of the feature point before movement is predicted from the movement result of the previous frame. Then, the feature points extracted from the image after the movement closest to the predicted position are associated with each other. By this method, the processing time of the self-localization has been decreased compared with the case where the feature point extraction and the feature amount description are used for the brute-force matching of the floor images before and after the movement. However, when a movement model with acceleration and deceleration is applied, it correlates with a feature point that should not be associated, resulting in the error of the measurement distance becoming large.

To solve this problem, a self-location method based on floor images considering AKAZE feature quantities and movement prediction is proposed. In the proposed method, matching according to feature quantities is implemented, which predicts the position of the moved floor image that the feature point extracted from the floor image before the movement to the current frame from the movement result of the immediately preceding frame. Based on the prediction result, the matching is calculated using the feature quantities described in the feature points extracted from the floor image before movement and the feature amount described in the feature point extracted from the floor image after the movement closest to the predicted position of the predicted after movement feature point. In this paper, we propose the visual odometry method to reduce computational cost compared with feature quantity matching using the brute force method.

In this paper, the proposed system is introduced in Sect. 2. The evaluation of the proposed method is given in Sect. 3. The paper is concluded in Sect. 4 with the summary and a discussion of future work.

II. PROPOSED SYSTEM

A. SYSTEM SUMMARY

For the estimation of movement, a monocular camera pointed perpendicularly toward the floor is installed. First, when the camera attached to the robot moves along the direction of the X axis parallel to the floor as shown in Fig. 1, the proposed system captures floor images continuously. Next, the proposed system detects keypoints extracted from the continuously captured two frames. Most keypoints detected from the floor are visible as an original vein and scratches. For the detection from each keypoint, accelerated KAZE features (AKAZE) proposed by Pablo [21] with the ability to detect descriptor robust to rotation and changing illumination is used. Then, the keypoints of the captured two frames are matched and the corresponding keypoints between the frames are estimated. The approach to increase the efficiency of matching is explained in detail in section 2.2. As shown in Fig. 1, when the camera faces perpendicularly toward the floor, the height between the floor and the camera is expressed as d, $1/2$ angle of the camera view, when the number of pixels in the x direction of the imaging sensor is w, and when the interest point (x^f, y^f) is moved to $(x^{f+1}, ^{f+1})$, the x and y movement quantities are v_x^{f+1} and v_y^{f+1}, respectively, and their relations can be express as follows.

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

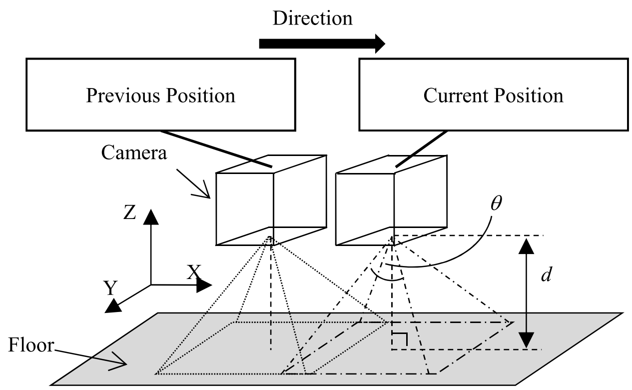

FIGURE 1. The proposed system measurement distance from floor images captured at the previous position and current position that it moved to following the directions from previous position. The proposed system measures distance by using the distance between the floor and CCD camera and matched each keypoint detected from captured pictures.

B. AKAZE FEATURE MATCHING WITH ESTIMATION METHOD

In this paper, to increase the efficiency of the keypoint matching of the image before and after movement from the captured previous frame and current frame, the movement quantity is estimated. Then, the coordinates of each keypoint movement are estimated.

From the descriptor matching of the captured images before and after movement, the calculation cost can be reduced by matching the keypoints on the predicted keypoints with the nearest keypoint of the next frame. The flowchart of the proposed system is shown in Fig. 2. Based on the keypoints determined before the movement, the position of the keypoints on the image after the movement is predicted. First, the floor image before and after the movement is captured using a monocular CCD camera positioned perpendicular to

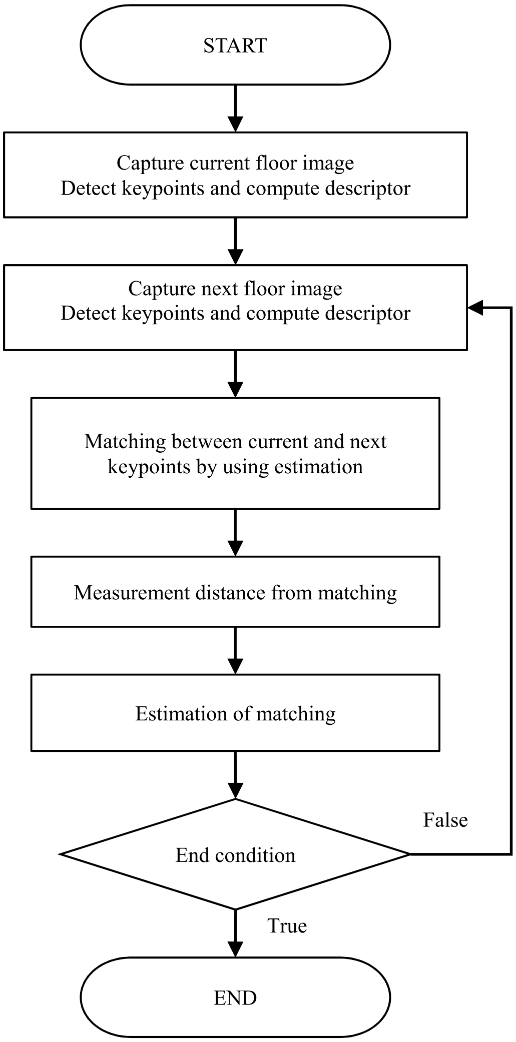

FIGURE 2. **Flowchart of proposed system. The proposed system has four steps. First, floor images captured at the current position and at the next position were detected keypoints and computed descriptors from accelerated KAZE features (AKAZE). Second, keypoints were matched by using Hamming distance and result of estimation. Third, the proposed method measures the distance from the result of the match. Finally, the proposed system estimates next position of the keypoint from the measured distance.**

the floor. Next, the extraction of the keypoints and the target of the descriptor are computed using the AKAZE method. The captured floor image is shown in Fig. 3, and the distribution of the keypoints extracted from the floor image is shown in Fig. 4. the keypoints are mainly from stains; the original vein is extracted, and the descriptor is then determined from the keypoints. Fig. 5 shows the results of feature matching in the proposed method. The proposed feature matching is shown in Fig. 6 with details. Based on the data acquired before, using the before and after movement descriptor points, the keypoints of the image are determined. The star symbol indicates one of the keypoints of interest extracted using AKAZE in the current floor image. The square symbol indicates the estimated position of keypoint predicted after the movement at current floor image. The prediction is made

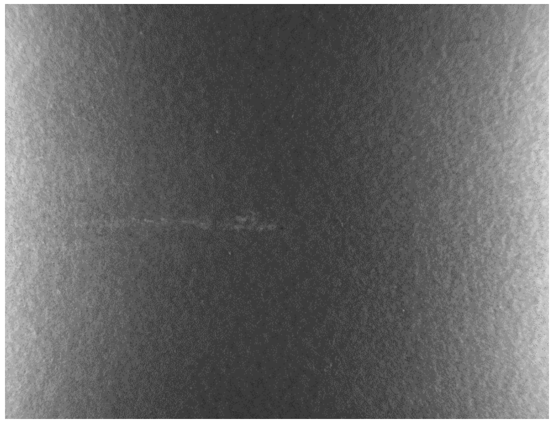

FIGURE 3. **Floor image captured by using a CCD camera. The CCD camera was installed 100 mm above the floor. Furthermore, the camera was covered with a light shielding plate. Two LEDs lights were installed 3 mm above the floor, illuminating toward the center of the image parallel to floor.**

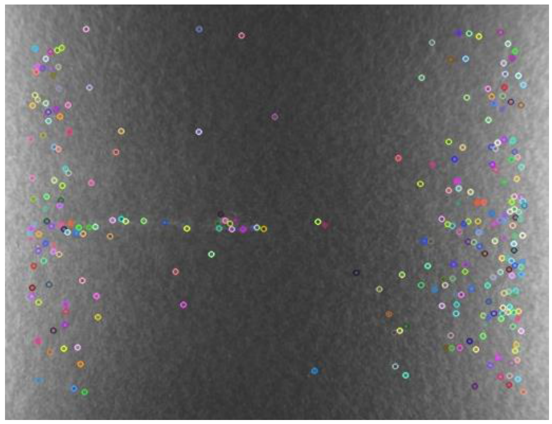

FIGURE 4. **Detected keypoints from Fig. 2. Most detected keypoints asperities and scratches on the floor.**

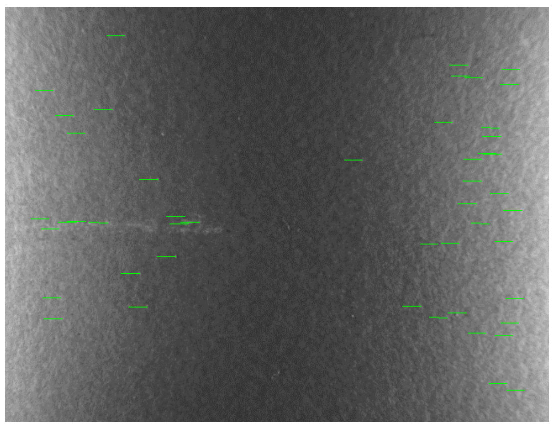

FIGURE 5. **The results of matching each keypoint after removing incorrectly related keypoints, as detected from floor images captured at current position and next position. For removing incorrectly related keypoints, standard deviation was used.**

assuming that the result of the self-localization performed in the previous frame will be the same in the next frame. The circle symbol with the number written shows the feature

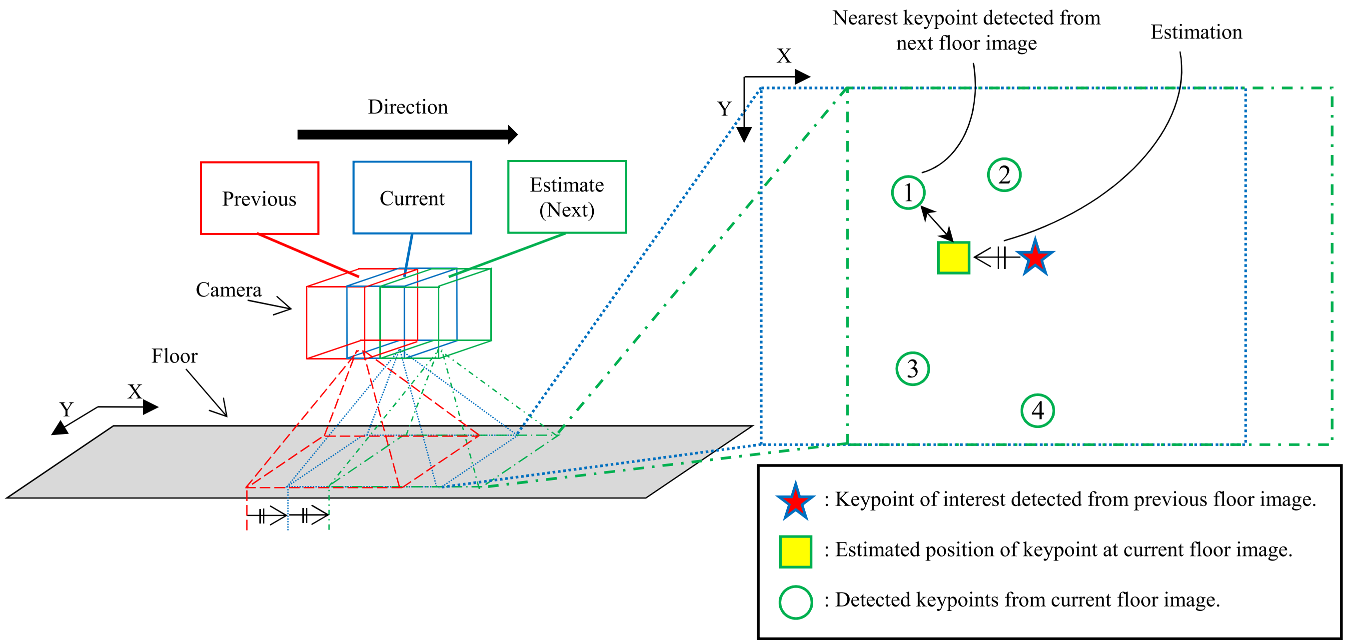

FIGURE 6. Matching algorithm for proposed system. For the matching, the proposed system has 3 steps. First, the estimated position was predicted by using the result of measurement distance from previous frame. Second, keypoint of interest detected from current floor image was matched to the nearest keypoint, which was measured between the estimated position of interest keypoint and keypoints which were detected from the next floor image. The proposed system decides the interest keypoint and the nearest keypoints to serve as paired keypoints from the Hamming distance.

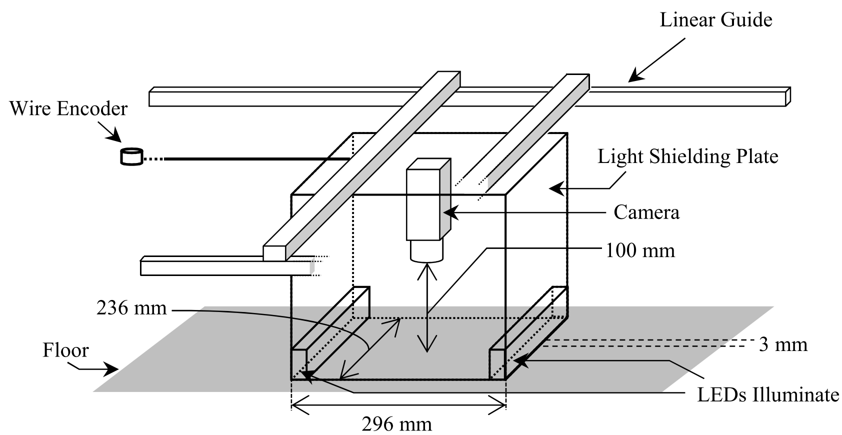

FIGURE 7. Experiment environment of proposed system. For the ground truth, a wire encoder was installed on the proposed system. Additionally, for moving linearly, a linear guide was installed.

points extracted from the floor image after movement using AKAZE. Here, the numbers sequentially show which feature points are close to the predicted point. For proposed feature quantity matching, from the keypoint extracted from the target point of the previous frame, the nearest movement position of the attention point is estimated, and the keypoints of the next frame are found sequentially. Next, from the ascertained keypoints and the estimated keypoints, the Hamming distance of their descriptor is calculated. Finally, a threshold is applied on the calculation result and the keypoints of the after-movement frame is determined. When the calculation does not meet the applied threshold, the next nearest Hamming distance of the descriptor is calculated and used as the current threshold. Moreover, from the result of matched keypoints obtained before and after movement, the incorrectly matched data is reduced by using standard deviation. The result of the matched keypoints is shown in Fig. 5. The keypoints acquired from the correspondence result are then used to estimate the current position. From the result of the estimated current position, the keypoints of the current frame are extracted as the possible coordinates of the next frame. For the movement position, the descriptor of the next frame determined from the descriptor of the current and previous frame is used.

The estimation of the descriptor of the current frame and next frame can be acquired when the camera capturing the image of the floor is under inertia; the camera moves gently parallel to the floor and when the frame rate is sufficiently fast. Observing from the X-axis, Y-axis, one frame before, current frame and the next frame, the camera recognizes the common aspects of the local feature. Correspondingly, when the local features position of the current frame is defined as $\left(x^{(f)}, y^{(f)}\right)$, the movement quantity of the current frame and the frame before in the X-axis and the Y-axis are $v_x^{(f)}$ and $v_y^{(f)}$ respectively, and when the angle of rotation during the rotary motion is defined as $\theta^{(f)}$, the keypoint's movement position of X-axis and Y-axis, $x^{'(f+1)}$ and $y^{'(f+1)}$ can be calculated from equation (2).

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

In this study, as the experiment only evaluates on the X-axis, the keypoint's movement position can be calculated using equation (3).

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

C. REMOVING INCORRECTLY RELATED KEYPOINTS

The proposed method sometimes corresponds to incorrect keypoints because some current frame keypoints may be absent in the next frame if the lighting has changed. Those incorrectly related points affected the measurement results. The proposed method uses standard deviation to remove incorrectly related points. First, calculate all X-axis and Y-axis movement using all related points from the current and next frame. Second, calculate the mean and standard deviation from all X-axis and Y-axis movement. Third, the average of X or Y-axis movement is defined as a_m, standard deviation of X or Y-axis as s_m and the measured movement using one related point as m. The proposed method can detect the correct related point from the following threshold of $a_m - s_m \leqq m \leqq a_m + s_m$. Finally, the average of X or Y-axis movement is defined as a_n, standard deviation of X or Y-axis as s_n and measured movement using one related point as n, and the proposed method can detect correct related point again from the following threshold of $a_n - s_n \leqq n \leqq a_n + s_n$. By doing so, the precision of the measurement result can be enhanced.

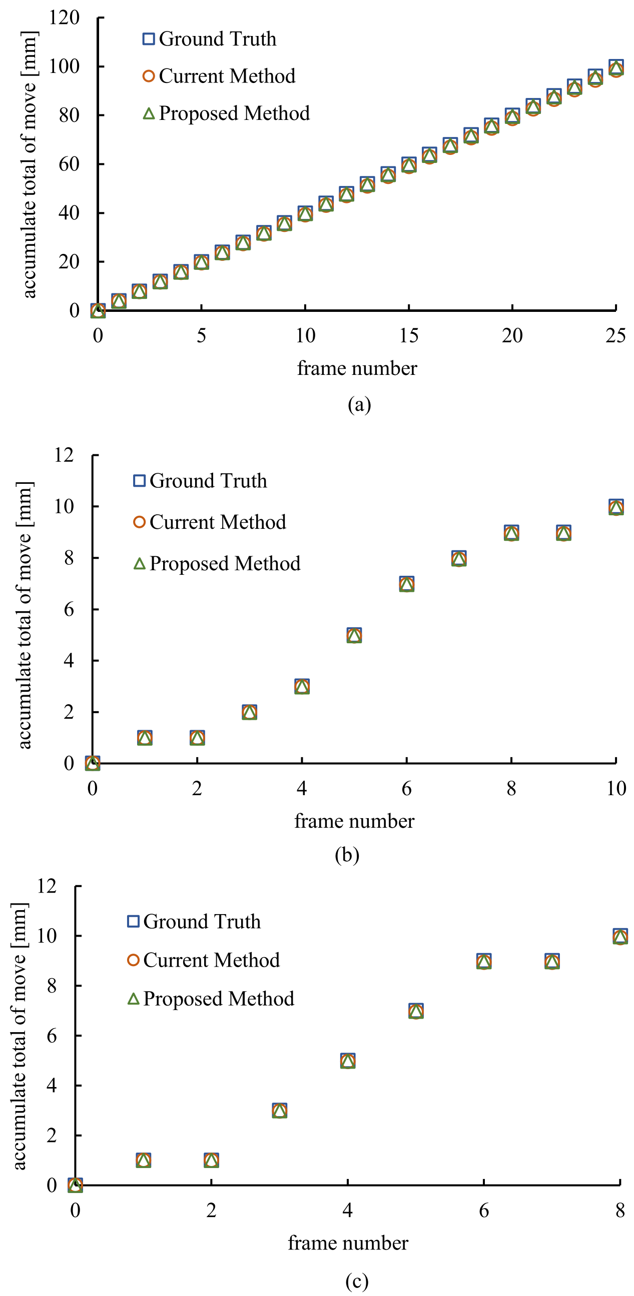

FIGURE 8. (a) – (c) For the test, equal speed model and acceleration model were tested. For acceleration model, 2 patterns of model were used. The result of measurement equal speed model and acceleration models by using the proposed method are shown in Fig. 7. For the test models, every model was tracked as ground truth.

III. EVALUATION EXPERIMENT

A. EXPERIMENTAL ENVIRONMENT AND METHOD

As shown in Fig. 7, for the experiment setup, STC-SB32POEHS (produced by OMRON SENTECH) was used for the CCD camera to capture the floor images. The size of the image is 648 × 494 pixels. In addition, MV-0614N (produced by NS-Lighting) was used for the lens is installed to CCD camera. The camera was installed facing downwards 100 mm vertically toward the floor. The proposed system covered with light shielding plate was installed with a 3-mm gap from the floor to avoid ambient light, which affects the captured images. For the illumination of floors and capturing original vein and scratches from the floor, two OPB-20015W2 (produced by OPTEX FA) were installed on both sides of the cover, near and parallel to the floor of the proposed system. P-72 (produced by TAJIMA ROOFING) was used as the observing object. In addition, linear guide was used for the proposed system to move linearly. The proposed system computes self-localization by using a PC (Intel Core

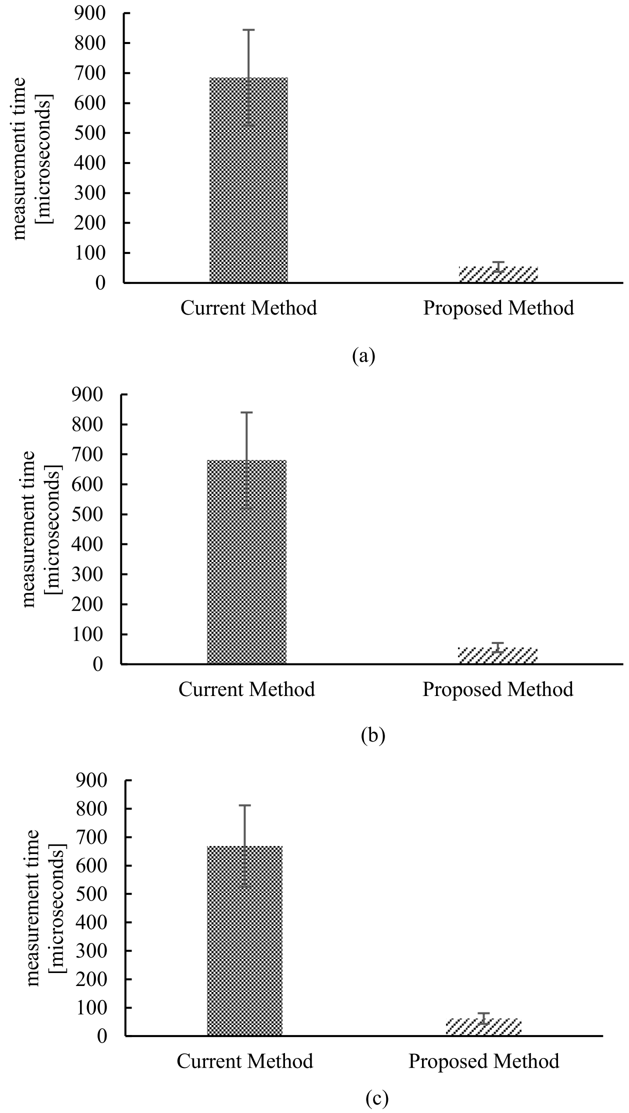

FIGURE 9. **(a) – (c) Results of computation time were measured using the proposed method. For the computation time, the proposed system was tested by using equal speed model 100 times. Furthermore, the proposed method was tested by using 2 accelerate models. Every accelerate model was tested 190 times. Fig. 8 shows the average of computation time.**

i7 7700K was used). For measuring the ground truth, a wire encoder (MLS-50-1080C-2000, produced by MICROTECH LABORATORY) was used. In the experiment, a comparison of the accuracy of self-localization and the processing time involved in matching was performed by the conventional method of comparing feature quantities with brute-force and by comparing the feature quantities in moving prediction and the proposed method of determining the order of feature matching. The true value was a measurement value obtained by the wire encoder. Therefore, a total of 200 floor images were captured with the wire encoder measuring 1 mm in the X-axis direction in advance using the proposed system. Using these images, we created one movement model of constant velocity motion and two movement models of acceleration and deceleration movement. The following three movement models are shown.

I) 100 mm by uniform speed model, 4 mm interval per frame
II) 10 mm by gain 1.25, sigmoid function model
III) 10 mm by gain 1.75, sigmoid function model

experiments were conducted while changing the image of the starting point of self-localization based on the three movement models. In the constant velocity movement model, a total of 100 image sets were created, and the proposed method of self-localization experiment was performed. In addition, in the movement model based on the sigmoid function that performs acceleration and deceleration, a total of 190 image sets were created and investigated by experiments. To compare processing time in matching, the processing time from the start to the end of the matching process was measured, and the average of the measurement time for the conventional method and the proposed method were calculated.

B. EXPERIMENTAL RESULT AND STUDY

The results of measurement for each model are shown in Fig. 8(a) to (c). The measurement time and standard deviation for each model are shown in Fig. 9(a) to (c). According to Fig. 8(a) to (c), both results of measurement distance are follow-ups to ground truth. In addition, according to Fig. 9(a) to (c), both results of computation time tried using the proposed method are faster than those tried by current method. As a result, the effectiveness of the proposed method is suggested.

IV. CONCLUSIONS

In this paper, a visual odometry method using AKAZE features with an estimate matching method was proposed. To estimate self-localization, the measurement of 2-D motion is necessary and important in various applications. For the measurement method, a contact system and noncontact system were proposed. The contact system measures self-localization by using a wheel encoder. As noncontact systems, methods using template matching or descriptors were proposed. However, these methods involve a high cost in estimation or are non-robust to illumination changing. To reduce computation time for the descriptor matching method, using the AKAZE method, which is robust to rotation imaging and illuminate changing, along with estimate descriptor matching method was proposed. In the realization of descriptor matching, brute-force matching was applied in the conventional method. In the proposed method, descriptor matching with movement estimation is used. This method estimates the position of the camera in the next frame by using movement during the previous and current frame. First, the proposed method searches the nearest point to target the keypoint's estimated position from keypoints detected in the next frame image. Second, the proposed system calculates the Hamming distance between the near keypoint's descriptor and target keypoint's descriptor. Third, for the matching, the system checks if Hamming distance is shorter than a threshold. Fourth, the system decides to match the detected keypoint and target keypoint or search for the next nearest point from the target keypoint. Finally, the proposed system calculates the distance between matched keypoints by using the standard deviation twice.

In the evaluation study, measurement distance and matching time were evaluated using the existing method and the proposed method. According to the experimental result, both results of measurement distance are the same with the ground truth. The proposed method works with high efficiency shows faster matching speed than the existing method.

Having presented the capabilities and limitations of the proposed system, it should be noted that when estimations are incorrect, the amount of matching increases; thus, the matching time is increased. Therefore, future work toward improving estimation is needed.

REFERENCES

[1] M. Dille, B. Grocholsky, and S. Singh, "Outdoor downward-facing optical flow odometry with commodity sensors," *Springer Tracts in Advanced Robotics*. Springer, 2010, pp. 183–193.

[2] I. Nagai and K. Watanabe, "Path tracking by a mobile robot equipped with only a downward facing camera," in *Proc. Int. Conf. Intell. Robots Syst.*, Sep. 2015, pp. 6053–6058.

[3] S. Chhaniyara, P. Bunnun, D. L. Seneviratne, and K. Althoefer, "Optical flow algorithm for velocity estimation of ground vehicles: A feasibility study," *Smart Sens. Intell. Syst.*, vol. 1, no. 1, pp. 246–268, Mar. 2008.

[4] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2008, pp. 3946–3952.

[5] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1150–1157.

[6] Y. Sakai, H. Lu, J.-K. Tan, and H. Kim, "Recognition of surrounding environment from electric wheelchair videos based on modified YOLOv2," *Future Gener. Comput. Syst.*, vol. 92, pp. 157–161, Mar. 2018.

[7] Y. Sakai, H. Lu, J. K. Tan, and H. Kim, "Environment recognition for electric wheelchair based on YOLOv2," in *Proc. ICBIP*, Aug. 2018, pp. 112–117.

[8] Y. Nakayama, H. Lu, Y. Li, and H. Kim, "Wide residual Networks for semantic segmentation," in *Proc. ICCAS*, Oct. 2018, pp. 1476–1480.

[9] M. Killpack, T. Deyle, C. Anderson, and C. C. Kemp, "Visual odometry and control for an omnidirectional mobile robot with a downward-facing camera," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 139–146.

[10] K.-C. Chen and W.-H. Tsai, "Vision-based autonomous vehicle guidance for indoor security patrolling by a SIFT-based vehicle-localization technique," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3261–3271, Sep. 2010.

[11] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[12] R. Mur-Artal and D. Juan Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Rob.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[13] S. Se, D. Lowe, and J. Little, "Vision-based mobile robot localization and mapping using scale-invariant features," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2001, pp. 2051–2058.

[14] B. Clipp, J. Lim, and J.-M. Frahm, "Parallel, real-time visual SLAM," in *Proc. Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 3961–3968.

[15] K. Konolige, M. Agrawal, and J. Solá, "Large scale visual odometry for rough terrain," *Rob. Res.*, vol. 66, pp. 201–212, Jan. 2011.

[16] T. D. Barfoot, "Online visual motion estimation using FastSLAM with SIFT features," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Aug. 2005, pp. 579–585.

[17] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "Svo: Semidirect visual odometry for monocular and multicamera systems," *IEEE Trans. Robot.*, vol. 33, no. 2, pp. 249–265, Apr. 2017.

[18] M. Persson, T. Piccini, M. Felsberg, and R. Mester, "Robust stereo visual odometry from monocular techniques," in *Proc. IEEE Intell. Vehicles Symp.*, Jul. 2015, pp. 686–691.

[19] J. Straub, S. Hilsenbeck, G. Schroth, R. Huitl, A. Möller, and E. Steinbach, "Fast relocalization for visual odometry using binary features," in *Proc. ICIP*, Sep. 2013, pp. 2548–2552.

[20] T. Morio, S. Aramaki, and S. Nakashima, "Visual odometry from floor images by using accelerated KAZE Features," in *Proc. Int. Conf. Ind. Appl. Engr.*, Mar. 2018, pp. 257–262.

[21] P. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *Proc. Br. Mach. Vis. Conf.*, Sep. 2013, pp. 13.1–13.11.

SHOTA NAKASHIMA (M'13) received the Ph.D. degree in electrical engineering from the Kyushu Institute of Technology, in 2010. From 2010 to 2012, he was an Assistant Professor with the Ube National College of Technology. He is currently a Senior Assistant Professor with Yamaguchi University. His current research interests include image processing, intelligent sensing, and artificial life. He is a member of the SCIE, IEEJ, and IIAE.

TOMOHIRO MORIO received the B.E. degree from Yamaguchi University, Yamaguchi, Japan, in 2017, where he is currently pursuing the M.E. degree with the Graduate School of Sciences and Technology for Innovation. His current research interests include measurement, image processing, intelligent sensing, and robotics.

SHENGLIN MU (M'11) received the B.E. degree in mechanical engineering and automation from Northeastern University, China, in 2007, and the M.E. and Ph.D. degrees in electronic and information systems engineering from Yamaguchi University, Japan, in 2010 and 2013, respectively. In 2013, he joined the Department of Electronic Control Engineering, National Institution of Technology, Hiroshima College. He is currently with the Graduate School of Science and Engineering, Ehime University. His current research interests include control engineering, especially intelligent control; mechatronics; robotics; and sensing technology. He is a member of the IEEJ and SICE.