

Calibrating Depth Sensors for Pedestrian Tracking Using a Robot as a Movable and Localized Landmark

Mitsuhiko Kimoto
IRC and Graduate School of Science and
Engineering,
ATR and Doshisha Univ.
Kyoto, Japan
kimoto2013@sil.doshisha.ac.jp

Masahiro Shiomi
IRC,
ATR.
Kyoto, Japan
m-shiomi@atr.jp

Takamasa Iio
Faculty of Engineering, Information and
Systems, and IRC
Tsukuba Univ and ATR.
Tsukuba and Kyoto, Japan
iio@iit.tsukuba.ac.jp

Katsunori Shimohara
Graduate School of Science and Engineering,
Doshisha Univ.
Kyoto, Japan
kshimoha@mail.doshisha.ac.jp

Norihiro Hagita
IRC
ATR
Kyoto, Japan
hagita@atr.jp

Abstract—Using multiple depth sensors enables us to accurately track pedestrians in real environments. Accurate pedestrian positions are essential for building an effective human-centered cyber world provided by location-based services. In particular, ceiling-mounted depth sensors can robustly track people in such environments. However, one important problem for this approach is the accurate calibration of the absolute sensor positions. This problem remains unsolved due to limited range and sensor distortions from a distance. Manual calibration is complicated and time-consuming, and the existing calibration method still has several limitations since it used a pedestrian as a movable landmark. Instead of a human landmark, we propose a method that uses a mobile robot as a movable and localized landmark to calibrate each sensor. We compared the calibration performance of the proposed and existing methods and showed that the former achieved more accurate calibration for both the absolute sensor and tracked pedestrian positions. Our proposed method with a mobile robot not only increased the accuracy of the calibration processes but also decreased human efforts.

Keywords—*person tracking, sensor calibration, mobile robot*

I. INTRODUCTION

Wide-area pedestrian tracking in real environments is critical for understanding human activities that provide essential knowledge to build an effective human-centered cyber world. Various techniques have been proposed, including sensor networks with multiple RGB cameras, such as laser range finders (LRFs) as the Hokuyo UTM series and the Velodyne LiDAR series, and such depth sensors as Microsoft Kinect and the ASUS Xtion series. One promising approach for wide-area pedestrian tracking with a multiple sensor network is using depth-based sensors on the ceiling, because this approach is robust to occlusions from the bodies of pedestrians. Moreover, due to using only distance-based information, privacy issues

are reduced compared to using RGB cameras [1]. In fact, this approach is broadly used to track pedestrians in the following real environments: for understanding the use of space by wide-area pedestrian tracking to analyze and anticipate their behaviors [2], planning approaches for pedestrian [3], providing such location-based services as distributing flyers [4], identifying a specific person using acceleration sensors [5] [6], and avoiding crowded situations due to a robot's existence [7].

One essential problem for such a ceiling-mounted depth sensor network system is its calibration, i.e., adjusting 6 DOFs (pan, tilt, roll, x, y and z) of each sensor to keep consistent observations between other sensors. However, common calibration methods that use shared observations between sensors are difficult to apply in ceiling-mounted depth sensors. To the best of our knowledge, only one method has been proposed that calibrates ceiling-mounted depth sensors for pedestrian tracking [8]. This study used a pedestrian as a movable landmark to acquire shared observations between sensors to calibrate relative/absolute sensor position relationships. But this decision, unfortunately, created several limitations even though it achieved good calibration accuracy; its limitations are scrutinized in the next section. Instead of employing a human landmark, we propose a method that uses a mobile robot as both a movable and a localized landmark for calibrating a sensor network (Fig. 1).

The remainder of this paper is organized as follows. Section II describes why the common calibration method cannot be applied to a ceiling-mounted depth sensor network system for pedestrian tracking and the strengths of our calibration method. Section III and IV describe our system settings for calibration and its mechanism. Section V describes the evaluations of our calibration method, and Sections VI and VII respectively describe our discussion and conclusion.

This research work was supported by JSPS KAKENHI Grant Numbers JP16K12505, 18H04121, 18H03311, and JP18J12552.

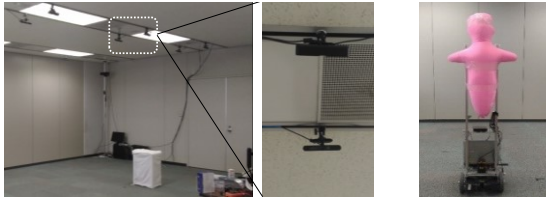


Fig. 1. Ceiling-mounted depth sensors for pedestrian tracking and a mobile robot as a movable and localized landmark for calibrating sensors

II. PROBLEM DEFINITION

A. Why Can't the Common Calibration Method be Applied for Ceiling-Mounted Depth Sensors?

A past study [8] thoroughly summarized the reasons for the difficulties of applying such common calibration methods as multiple video camera calibrations [9-11] in ceiling-mounted depth sensors. We briefly explain them in this section by following [8].

One basic approach for calibrating depth sensors is point cloud matching, which uses static objects in the background between sensors and resembles approaches like SLAM [12, 13]. There are two main reasons why this approach cannot be used in ceiling-mounted depth sensors: distorted point cloud information and the difficulties of identifying the shared features among the sensors. The former reason indicates that since ceiling-mounted sensors observe static objects beyond their nominal range, many measurements are missing, and range distortion occurs [14]. This phenomenon also complicates extracting a floor plane, which calibrates the pitch, the roll, and the height of the sensors. The latter reason indicates that due to a narrow field of view from ceiling-mounted sensors, identifying the features shared among them is often difficult. This problem also complicates marker-based calibration, which is another common approach to this problem [15], because many markers are necessary to achieve good calibration for ceiling-mounted sensors with narrow sensing areas.

Based on these reasons, Glas et al. proposed a calibration method that used the observations of a human landmark for sensor calibration by integrating LRF and depth sensors [8], because the observations of a human head are typically within the nominal range of ceiling-mounted depth sensors and thus do not suffer from distortion. Furthermore, using LRF sensors for calibration enables systems to easily find shared observations among sensors. Even though their study accurately calibrates ceiling-mounted depth sensors, some limitations exist. In the next subsection we describe these problems and describe our approach for solving them.

B. Limitations of Existing Method for Calibrating Ceiling-Mounted Depth Sensor and Our Approach

As described above, the past study proposed a method for calibrating ceiling-mounted depth sensors by integrating LRF and depth sensors. However, its method has several limitations. In this section we describe the problems caused by them.

1) *A person needs to walk around to be a movable landmark:* The past study employed a person as a movable landmark to create shared observations among sensors.

However, it needs additional human resources for calibration, especially in wide sensing environments. For example, a past study covered about 900 m² using more than 50 depth sensors, but walking this entire environment would be quite strenuous [1].

2) *Unstable height due to walking:* This problem is related to using a human landmark. The past study used fixed height information to calibrate the sensor's pitch, roll, and height by assuming a person's height does not change during calibration. But when a person is walking, her height actually does change. Of course, since it is difficult for people to maintain the same height while walking, this problem is unavoidable with a human landmark.

3) *Must overlap observation areas of the sensors:* The past study assumed that the coverage area of at least two sensors will overlap due to shared observations. In other words, its approach failed to calibrate a sensor that does not overlap its sensing area with other sensors. To cover this problem, it proposed using an LRF to make an overlapped sensing area, but this step also increases the total burden to prepare a calibration system.

4) *Absolute sensor position calibration:* The past study used the shared observations among sensors to calibrate their positions, but the accuracy of their absolute position calibration was strongly influenced by their initial settings. Because the shared observations from the human landmark are the relative position information based on the initial sensor positions, if the initial positions are shifted, the calibration results for the absolute sensor position also shifted.

5) *Our approach:* As described above, the past study's limitations are based on its use of a human landmark. In this study, we solved these limitations with a mobile robot as a movable and localized landmark. This approach decrease human's efforts to walk around a sensing area as a human landmark for a calibration procedure and provides to the calibration system accurate localized position with stable height information while moving. Moreover, using a movable and localized landmark enables the system to calibrate a sensor that does not overlap its sensing area with others, because the robot provides global reference *points* for each sensor by moving around. The next section describes the details of our calibration system design.

III. SYSTEM DESIGN

Figure 2 shows an overview of our entire system, which consists of depth sensors, a mobile robot, a human-tracking system, and a calibration system.

A. Depth Sensors

We installed 10 ASUS Xtion Pro Live 3D range sensors as depth sensors for pedestrian tracking in our experiment environment (Fig. 3). The sensors were mounted in rows on the ceiling at an average height of 2.6 m, covering a 8.2 m long by 4.5 m wide tracking area. They were attached to the ceiling upside-down in rows in alternate directions to minimize interference and maximize coverage.

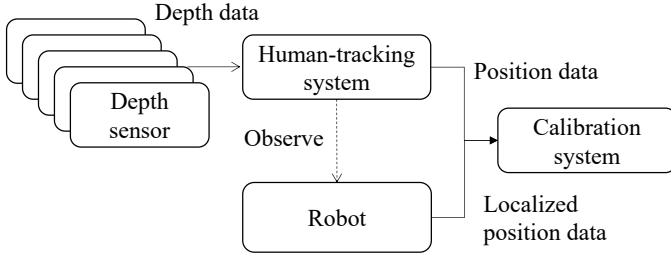


Fig. 2. Overview of our calibration system

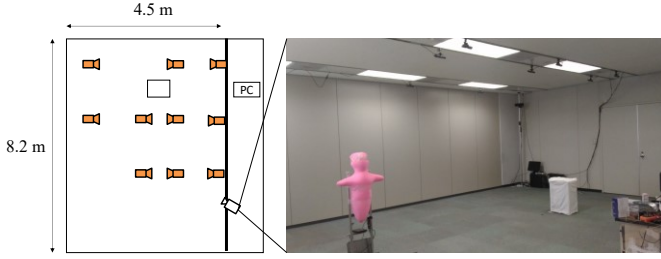


Fig. 3. Sensor arrangement and environment in our study

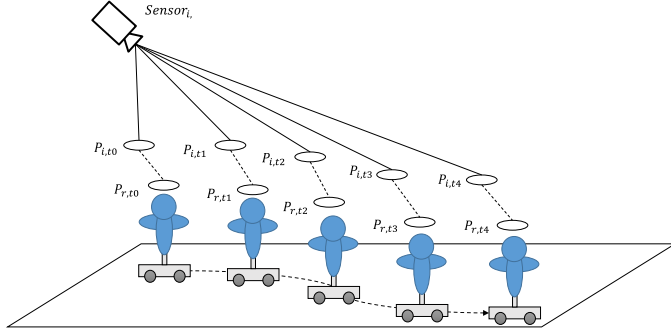


Fig. 4. Estimated positions $P_{i,t}$ from the sensor i and the robot's positions $P_{r,t}$.

B. Human-tracking and calibration systems

We captured the data from depth sensors by five desktop PC's with up to four sensors connected to each PC. These sensor data were streamed over a wired network to a Core i5 PC running tracking software written in Java [1]. In this study, these sensors were used to detect the tops of people's heads, not for full-body skeleton tracking (see a previous work for details of the tracking algorithm [1]). For this purpose, the sensor angles were approximately 30-60 degrees from a horizontal degree. The initial sensor positions were calibrated by a human.

The calibration system was also written in Java and operated on the same PC. The tracked position data and the localized robot's position data were sent to the software for calibration through a wired/wireless network. For documentation and repeatability of the results, we conducted offline processing, but the software can be used online with live data.

C. Mobile Robot as a Movable and Localized Landmark

Figure 1 shows our mobile robot, which has a human-sized head (at 135 cm height) and its shoulders are tracked as an entity by the tracking system. We used an iCart-mini robot as a

mobile base, which is controlled using a Robot Operation System (ROS) framework. For its localization, we used the Adaptive Monte Carlo Localization (AMCL) method [16], which is a common probabilistic localization algorithm for ROS-based mobile robots. In our experimental environment, we made a map for localization beforehand, and the measured average localization accuracy was in the order of 5 cm.

IV. CALIBRATION MECHANISM

A. Overview

Since we used the robot's localized position, we can calibrate each degree of each sensor one by one, unlike the past study. In other words, using a mobile robot as a movable and localized landmark, we can directly calibrate the absolute positions and the degrees for each sensor. First, the system calibrates the pitch, the roll, and z so all the sensors are coplanar and then calibrates the yaw, x , and y for all sensors to adjust their absolute positions.

B. Pitch-Roll-Z Calibration

For calibrating each sensor, the system first processes pitch (θ) and roll (ψ) to fit between the planes of the head part observed from the sensors and localized from the robot. For the pitch and roll calibration, the system conducts a grid search to find better values that minimize the standard deviations between the observed z positions from the sensors and the robot's z positions through all the data sets. By this process, the head plane coincides with the height level of the robot's position (ground-truth) plane. Then the system shifts the z positions to match the observed heights with the robot's height. We defined the set positions of sensor i as

$$Pos(i) = \{P_{i,0} \dots P_{i,t}\}, \quad (1)$$

where $P_{i,t}$ is the tracked position from sensor i at time t (Fig. 4). For a grid search of the pitch and the roll, the system calculates the standard deviation of the Euclidean distance of the z axis between the translated tracked positions and robot position $Pos(r)$. We used a rigid transformation matrix, $T_{\theta,\psi}$, which rotates the observed position from the sensors with specified pitch and roll values to calculate the following standard deviations:

$$Calibrate(\theta_i, \psi_i) = \underset{\theta \in \{\theta_{i,min} \dots \theta_{i,max}\}, \psi \in \{\psi_{i,min} \dots \psi_{i,max}\}}{\operatorname{argmin}} ZDev(Pos(i), \theta, \psi) \quad (2)$$

$$ZDev(Pos(i), \theta, \psi) = STDEV_z(T_{\theta,\psi} \cdot Pos(i), Pos(r)), \quad (3)$$

where θ_{min} and θ_{max} , which indicate the searching range for sensor i , are between ± 60 degrees from the initial θ of sensor i with 0.01 intervals, ψ_{min} and ψ_{max} , which indicate a searching range for sensor i , are between ± 60 degrees from the initial ψ of sensor i with 0.01 intervals, and $STDEV_z$ calculates the standard deviation of the Euclidean distance of the z axis between two position data sets.

After calibrating the pitch and the role of each sensor, we conducted a grid search for the z positions of each sensor that minimized the Euclidean distance of the z axis using rigid translation matrix T_z , which translates the observed position from the sensors in the z axis by considering of $T_{\theta,\psi}$, and robot's position $Pos(r)$ as follows:

$$\text{Calibrate}(z_i) = \underset{z \in \{z_{i,\min}, \dots, z_{i,\max}\}}{\text{argmin}} \text{ZDist}(\text{Pos}(i), z) \quad (4)$$

$$\text{ZDist}(\text{Pos}(i), z) = \text{Dist}_z(\mathbf{T}_z \cdot \text{Pos}(i), \text{Pos}(r)), \quad (5)$$

where z_{\min} and z_{\max} indicate a searching range for sensor i between ± 1000 mm from the initial z position of sensor i at 5 mm intervals and Dist_z calculates the Euclidean distance of the z axis between the position data sets.

C. Yaw-X-Y Calibration

After calibrating the pitch, the roll, and z for each sensor, the system processes the yaw (ϕ) and then again calibrates the x and y positions by a grid search. For a grid search of the yaw, similar to the pitch and roll calibration, the system calculates the standard deviation of the Euclidean distance of the x and y axes between the translated tracked positions with rigid transformation matrix \mathbf{T}_ϕ , which rotates the observed position from the sensors with specified yaw values, and the robot's position $\text{Pos}(r)$ as follows:

$$\text{Calibrate}(\phi_i) = \underset{\phi \in \{\phi_{i,\min}, \dots, \phi_{i,\max}\}}{\text{argmin}} \text{XYDev}(\text{Pos}(i), \phi) \quad (6)$$

$$\text{XYDev}(\text{Pos}(i), \phi) = \text{STDEV}_{x,y}(\mathbf{T}_\phi \cdot \text{Pos}(i), \text{Pos}(r)), \quad (7)$$

where ϕ_{\min} and ϕ_{\max} indicate a searching range for sensor i between ± 60 degrees from the initial ϕ of sensor i with 0.01 intervals and $\text{STDEV}_{x,y}$ calculates the standard deviation of the Euclidean distance of the x and y axes between the two position data sets.

After calibrating the yaw of each sensor, we conducted a grid search for the x and y positions of each sensor that minimized the Euclidean distances of the x and y axes using rigid translation matrix $\mathbf{T}_{x,y}$, which translates the observed position from the sensors in the x and y axes and robot's position $\text{Pos}(r)$ as follows:

$$\text{Calibrate}(x_i, y_i) = \underset{x \in \{x_{i,\min}, \dots, x_{i,\max}\}, y \in \{y_{i,\min}, \dots, y_{i,\max}\}}{\text{argmin}} \text{XYDist}(\text{Pos}(i), x, y) \quad (8)$$

$$\text{XYDist}(\text{Pos}(i), x, y) = \text{Dist}_{x,y}(\mathbf{T}_{x,y} \cdot \text{Pos}(i), \text{Pos}(r)), \quad (9)$$

where x_{\min} and x_{\max} , which indicate a searching range for sensor i , are between ± 1000 mm from the initial x position of sensor i with 5 mm intervals, y_{\min} and y_{\max} indicate a searching range for sensor i between ± 1000 mm from the initial y position of sensor i at 5 mm, and $\text{Dist}_{x,y}$ calculates a Euclidean distance of the x and y axes between the position data sets.

V. EVALUATION

A. Procedure

Following the past study, we investigated the performance of our proposed calibration system based on two perspectives: the sensor's absolute positions and the detected head positions. For each of these evaluation measurements, we compared the performances of our proposed method and the alternative method that is previously proposed approach [8].

We only used depth sensors for the calibration in this study (no LRFs), and used our mobile robot to record position data for evaluating both methods with offline processing as described above, to clearly compare the performance of the

two methods by using robot's localized position as ground-truth. The robot needs six minutes to cover all sensing area.

B. Sensor Position Accuracy

We again followed the past study's procedure [8] for sensor position accuracy: a laser range measurement device (Leica Geosystems, Leica DISTO X310) that measured the location of each sensor in an (x, y, z) coordinate system relative to the room's walls. In this study, the initial sensor positions for the tracking system were calibrated by a human experimenter with adequate experience with manual calibration tasks, without knowledge about actual measurement information. The calibrated sensor positions have 369.39 mm error in average compared to the measured sensor positions. The sensor positions are well calibrated from a relative position relationship perspective and enable a human-tracking system to robustly track people. But since these absolute positions were slightly moved from the actual sensor positions, a certain amount of shifted positions is directly influenced by the above position errors. To calibrate the sensor positions, our mobile robot moved around the environment once, and its data was used to calibrate sensor positions with both methods.

C. Head-Tracking Accuracy

Next, we used both the calibrated sensor position data from each method to track the head positions. For this purpose, we again tracked our mobile robot by our system; the robot moved around the same route twice. We used the robot's localized position as the ground-truth positions for the performance evaluations, which is another reason why we did not use a human landmark for the alternative method. We gathered a total of about 4000 paired sensor data for both trials, which are sufficient for evaluation purposes because since the past study used 360 measurement points at specific positions only.

D. Results

Fig. 5 shows the robot's localized positions (a), the observed positions from the tracking system (b) and integrated them (c), note that the sensor positions for the tracking system were manually calibrated by a human experimenter. Each sensor's observations plotted with different colors in the Fig. 5.

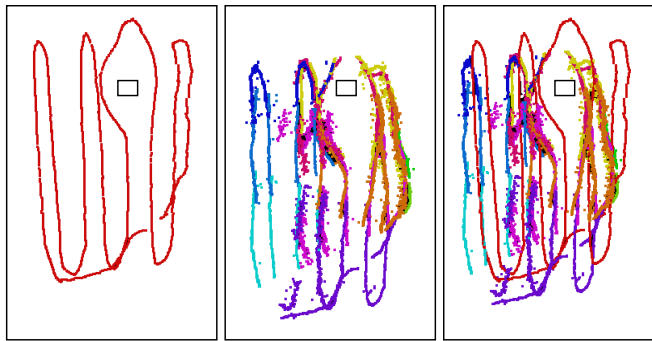
Fig. 6 shows the integrated trajectories with the calibrated sensor positions by the proposed method (a) and the alternative method (b). All of the figures showed that the shapes of the observed trajectories from the tracking systems resembled the robot's trajectories, but the absolute positions were shifted more than the robot trajectories without the proposed method.

Table I shows the evaluation results of the sensor position calibration and the head-tracking accuracies. We computed the root-mean-squared (RMS) error in x , y , z and combined them. The results of the sensor position showed that the proposed method achieved 100.17 mm error, and the alternative method achieved 346.97 mm error on average. The larger errors in the alternative method were caused by its initial sensor positions, which were calibrated by a human experimenter. The relative position relationship between the sensors was well calibrated, but the absolute sensor positions slightly shifted compared to the real sensor positions, as described above. Therefore its position errors influenced the alternative method's performance. On the other hand, our proposed method

calibrated the absolute sensor positions well regardless of the initial sensor positions, because it used the localized position from the mobile robot system. Our proposed method showed better accuracy than the alternative method.

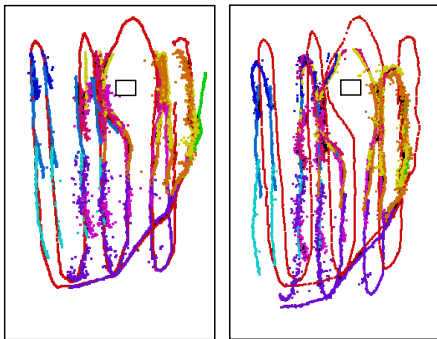
The results of the tracked human positions showed that the proposed method achieved 135.29 mm error, and the alternative method achieved 694.04 mm error in average. The larger error of the tracking positions was also caused by the initial sensor positions. Even though the human-tracking system tracked the people well from relative position relationships, the absolute positions were different than the actual positions. Again, the proposed method tracked the robot's position well regardless of the errors of the initial positions of the sensors.

Compared to the past study's evaluation results [8], in our setting, the performance of the alternative method is slightly inaccurate from the original results for reasons based on its sensors' initial positions and the small overlapping sensing area among the sensors. Our proposed method achieved similar performance with the past study's evaluation results using both depth and LRF sensors, and this result also indicates its promising performance.



(a) Robot's positions (b) Observed positions (c) Integrated both positions

Fig. 5. Robot's positions and observed positions from tracking system using human-calibrated sensor positions.



(a) Proposed method (b) Alternative method

Fig. 6. Robot's positions and observed positions from tracking system

TABLE I. EVALUATION RESULTS ABOUT SENSOR POSITION AND POSITION TRACKING ACCURACIES

	Alternative	Proposed
Sensor position accuracy (RMS, mm)	346.96	100.17
Position tracking accuracy (RMS, mm)	694.04	135.29

TABLE II. ADDITIONAL EVALUATION RESULTS

	Alternative	Proposed
Sensor position accuracy (RMS, mm)	379.53	115.91

E. Additional Evaluation with Non-overlapped Sensors

The proposed method showed the advantages of the calibration performance compared to the past study. In this subsection, we conducted an additional evaluation to calibrate sensors whose sensing areas do not overlap with other sensors to show different advantages of our proposed method compared to the alternative method. With a mobile robot as a landmark, each sensor can get global reference points from the robot's localized results, and this position information can be used for calibration without shared observations with other depth sensors.

In this evaluation, we only used five depth sensors in the middle of the room to create a sparse sensing area and again calibrated these sensor positions with the alternative and proposed methods. We again used our mobile robot to be tracked by the system, and it moved around for six minutes twice on the same route.

The performance of this evaluation with the sparse data set is shown in Table II. The proposed method's performance is slightly incorrect (about 15 mm) compared to the performance that used all of the sensors. But the alternative approach's performance was twice as bad (about 30 mm). These results indicate that using a mobile robot to provide global reference points is more robustness for calibrating absolute sensor positions than just using relative reference points.

VI. DISCUSSION

A. Implications

The evaluation results showed the advantages of using a mobile robot as a movable and localized landmark instead of a pedestrian landmark. The calibrated sensor positions and tracking results through our proposed method outperformed the alternative method's results, and the proposed method decreased the human efforts to gather sensor data. These advantages will be effective with large sensing areas, such as a sensor network system that covers an entire shopping mall.

One possible future work will integrate both an environmental human-tracking system and an onboard human-tracking system of a robot. If the robot's tracking system estimates the people's position during calibration, it can be used as different global reference points to increase the

accuracy of the calibration processes. From another perspective, if the robot can change its height and send its information to the calibration system, z-axis calibration will be more accurate.

B. Possible Applications

We used a mobile robot to provide global reference points for calibrating a human-tracking system. We believe that such a localized landmark can be used for calibrating different kinds of systems, such as Wi-Fi or any wireless signal-based systems. In fact, a past study previously tried to automatically calibrate an ultra-wide-band-tracking system with a mobile robot [17]. Mobile robots can be used not only as movable logger devices but also as a tool to calibrate several kinds of sensor positions.

C. Limitations

This study has several limitations. Since it was conducted with our sensing environment and a mobile robot, we need to investigate our proposed system's performance in different environments. In particular, in a large environment, the robot's localization error will increase and influence the accuracy of the calibration procedures. Moreover, we used a grid search approach to calibrate each sensor due to its simplicity and to avoid local minimum problems, but we did not test other methods such as MCMC that may realize fast calculation. The accuracy of the map information for localizing the robot position also influences the accuracy of the calibration processes. However, we believe that our setting offers essential knowledge for researchers who are interested in autonomous calibration by ceiling-mounted depth sensors.

VII. CONCLUSION

We proposed a method that calibrates with ceiling-mounted depth sensors using a mobile robot as a movable and localized landmark. This approach enables us to calibrate a human-tracking system without additional human resources and increases the accuracy of the sensor arrangement, which is directly related to the usefulness of the system and the accuracy of the tracking positions. Our proposed method calibrates each degree and the position of each sensor by directly comparing the paired positions between the robot's localized position and the detected robot's position by the tracking system. The system calibrated them by minimizing the deviation of the Euclidean distance to calibrate the pitch, the roll, and the yaw, and minimizing the Euclidean distance of x-y-z for calibrating the x, y, and z positions.

The proposed method achieved higher accuracy than the past study from two viewpoints: sensor position accuracy and the tracking performance of the head part. The evaluation results showed that the proposed method calibrated the sensor positions regardless of the initial sensor position errors, unlike the past study's method. An additional evaluation using sparse sensor arrangements showed that the proposed system calibrated sensors that did not overlap with other depth sensors with a robot's landmark. These evaluation results showed the potential of a calibration system that uses a mobile robot.

REFERENCES

- [1] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person Tracking in Large Public Spaces Using 3-D Range Sensors," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522-534, 2013.
- [2] T. Kanda, D. F. Glas, M. Shiomi, and N. Hagita, "Abstracting people's trajectories for social robots to proactively approach customers," *IEEE Transactions on Robotics*, vol. 25, no. 6, pp. 1382-1396, 2009.
- [3] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "A robot that approaches pedestrians," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 508-524, 2013.
- [4] C. Shi, M. Shiomi, C. Smith, T. Kanda, and H. Ishiguro, "A Model of Distributional Handing Interaction for a Mobile Robot," in *Robotics: Science and Systems*, pp., 2013.
- [5] M. Shiomi and N. Hagita, "Finding a person with a wearable acceleration sensor using a 3D position tracking system in daily environments," *Advanced Robotics*, vol. 29, no. 23, pp. 1563-1574, 2015.
- [6] C. Shi, S. Satake, T. Kanda, and H. Ishiguro, "How Would Store Managers Employ Social Robots?," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, Christchurch, New Zealand, pp. 519-520, 2016.
- [7] H. Kidokoro, T. Kanda, D. Brscic, and M. Shiomi, "Simulation-Based Behavior Planning to Prevent Congestion of Pedestrians Around a Robot," *Robotics, IEEE Transactions on*, vol. 31, no. 6, pp. 1419-1431, 2015.
- [8] D. F. Glas, D. Bršičič, T. Miyashita, and N. Hagita, "SNAPCAT-3D: Calibrating networks of 3D range sensors for pedestrian tracking," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 712-719, 2015.
- [9] A. W. Senior, A. Hampapur, and M. Lu, "Acquiring multi-scale images by pan-tilt-zoom control and automatic multi-camera calibration," in *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, pp. 433-438, 2005.
- [10] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *Computer*, vol. 34, no. 8, pp. 57-66, 2001.
- [11] P. Natarajan, P. K. Atrey, and M. Kankanhalli, "Multi-Camera Coordination and Control in Surveillance Systems: A Survey," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, no. 4, pp. 1-30, 2015.
- [12] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D Point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927-941, 2008.
- [13] N. Mohsin and S. Payandeh, "Localization and identification of body extremities based on data from multiple depth sensors," in *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on*, pp. 2736-2741, 2017.
- [14] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437-1454, 2012.
- [15] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *JOSA A*, vol. 4, no. 4, pp. 629-642, 1987.
- [16] F. Dellaert, D. Fox, W. Burgard, and S. Thrun, "Monte carlo localization for mobile robots," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, pp. 1322-1328, 1999.
- [17] A. Canepa, Z. Talebpour, and A. Martinoli, "Automatic Calibration of Ultra Wide Band Tracking Systems Using A Mobile Robot: A Person Localization Case-study," in *The International Conference on Indoor Positioning and Indoor Navigation (IPIN 2017)*, pp., 2017.