IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2015; **10**(s1): S108–S115
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.22171

Paper

Calibration of Multiple Kinects with Little Overlap Regions

Mitsuru Nakazawa*[a], Non-member
Ikuhisa Mitsugami*, Non-member
Hitoshi Habe**, Non-member
Hirotake Yamazoe***, Non-member
Yasushi Yagi*, Non-member

When using multiple Kinects, there must be enough distances among neighboring Kinects to avoid spoiled range data caused by the interference of their infrared speckle patterns. In the arrangement, their overlapped regions are too small to apply existing calibration methods using correspondences between their observations straightforwardly. Therefore, we propose a method to calibrate Kinects without large overlapped regions. In our method, first, we add extra RGB cameras in an environment to compensate overlapped regions. Thanks to them, we can estimate their camera parameters by obtaining correspondences between color images. Next, for accurate calibration, which considers range data as well as color images of Kinects, we optimize the estimated parameters by minimizing both the errors of correspondences between color images and those of range data of planar regions, which exist in a general environment such as walls and floors. Although our method consists of conventional techniques, its combination is optimized to achieve the calibration. © 2015 Institute of Electrical Engineers of Japan. Published by John Wiley & Sons, Inc.

Keywords: Kinect, calibration, interference

Received 16 August 2014; Revised 10 April 2015

1. Introduction

Recently, Kinect [1] and other types of consumer range sensors have attracted the attention of researchers, because they can simultaneously capture both metric range data and color images at low cost. Using a Kinect, many applications have been developed such as people detection [2], gait identification [3], pose estimation [4,5], and 3D shape reconstruction [6,7]. If we could utilize multiple Kinects for these applications, it would be possible to enlarge the area that these applications cover. To realize a system using multiple Kinects, it is one of fundamental and essential tasks to calibrate their camera parameters, that is, intrinsic parameters including their focal length and radial distortion and extrinsic ones consisting of their rotation and translation.

For RGB cameras, some methods to calibrate their intrinsic and extrinsic parameters have been proposed. For their intrinsic calibration, there exist methods such as the one proposed by Zhang [8]. The existing methods can also be used for the intrinsic calibration of Kinects straightforwardly. On the other hand, for the extrinsic calibration of RGB cameras, there exist methods that are based on corresponding points between their images [9]. In these methods, it is desirable that RGB cameras have large overlapping regions to acquire a sufficient number of corresponding points between their images. In contrast, for multiple Kinects, it is not desirable because when Kinects have large overlapping regions, their infrared (IR) speckle patterns, which are projected to construct range data, interfere with neighboring Kinects. The interference causes missing areas in the resultant range data of each Kinect. This is referred to as the "interference problem." Considering the interference problem, there must be sufficient distance between neighboring Kinects. In this arrangement, the existing method cannot be straightforwardly applied for their extrinsic calibration because it is impossible to obtain a sufficient number of correspondences between color images from their small overlapping image regions. Moreover, for their accurate extrinsic calibration, it is necessary to consider not only color images but also range data because the two captured data of Kinects involve their independent errors.

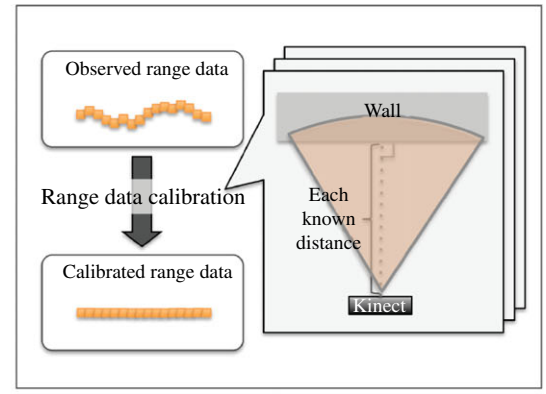
Therefore, we propose a calibration method for multiple Kinects whose overlapping regions are not large enough, considering not only the errors of color images but also those of range data. In extrinsic calibration of our method, first, we add extra RGB cameras in an environment to solve the shortage of the overlapped regions. In the overlapped regions among Kinects and extra RGB cameras, we can obtain a sufficient number of corresponding points between color images. They enable the estimation of their camera parameters by an existing method. Next, for accurate calibration, which considers not only the errors of color images but also those of range data, we optimize the estimated parameters by minimizing both the errors of correspondences between color images and those of range data of planar regions, which exist in a general environment such as walls and floors. Although our method is composed of well-known processes, we properly select and combine them with the consideration of the interference problem, which existing methods for calibrating extrinsic parameters of multiple Kinects [10–14] do not consider.

2. Accurate and Efficient Kinect Calibration

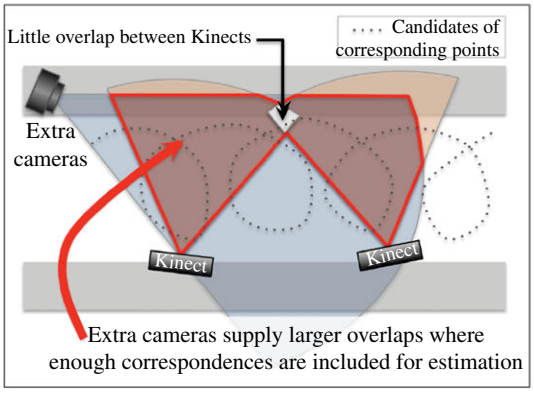
Inside a Kinect, there are two types of cameras, that is, an RGB camera and an IR camera that observes speckle patterns projected by an IR projector to construct range data. However, it can be

[a] Correspondence to: Mitsuru Nakazawa. E-mail: mnacazawa@gmail.com

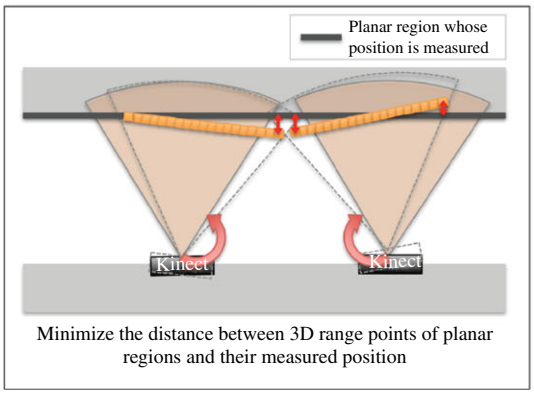
* The Institute of Science and Industrial Research, Osaka University, 8-1, Mihogaoka, Ibaraki, Osaka 567–0047, Japan
** Faculty of Science and Engineering, Kinki University, 3-4-1 Kowakae, Higashiosaka, Osaka 577–8502, Japan
*** Osaka School of International Public Policy, Osaka University, 1-31 Machikaneyama, Toyonaka, Osaka 560-0043, Japan



(a)



Step 1: Initial estimation using correspondences supplied by extra RGB cameras



Step 2: Optimization additionally using range data of planar regions whose position is measured

(b)

Fig. 1. Outline of the proposed calibration method. (a) Intrinsic parameter calibration and (b) Extrinsic parameter calibration

regarded that they have the same intrinsic and extrinsic parameters through alignment between the range data and the color image, which is provided by the Kinect SDK function with calibration information on factory default configuration. Under this premise, we discuss how to calibrate the intrinsic and extrinsic parameters of multiple Kinects.

As intrinsic parameters of a Kinect, we consider the following two kinds of parameters. The first ones are intrinsic camera parameters including the focal length, the image center, and the radial distortion. These parameters can be calibrated by conventional methods such as [8] straightforwardly. The second ones are parameters to compensate range errors, which increase according to the object distance of a Kinect [15] (Fig. 1(a)). These parameters are estimated using range data that are captured under the condition that the true range values can be measured.

On the other hand, extrinsic parameters of a Kinect consist of the rotation and translation as generally defined. When multiple Kinects are located, the distances among neighboring Kinects must be wide enough to avoid the interference problem. To calibrate extrinsic parameters of Kinects that have few overlapping regions, we have to consider a feasible solution method. One approach to calibrate such Kinects is to align each Kinect to 3D world coordinates independently. This approach can achieve extrinsic calibration of Kinects by using 2D points captured from color images [16] or 3D points captured from range data [17]. However, this approach is impractical because the larger the number of Kinects is, the more it is required to spend time and effort for the alignment of each Kinect. Another approach is to first estimate relative rotation and translation of Kinects by using the correspondences between captured data of Kinects, and then align them to 3D world coordinates. This approach is more practical than the former for the extrinsic calibration of many Kinects. However, it is impossible to apply the approach to our case straightforwardly because their small overlapping regions cannot provide a sufficient number of correspondences. In the past, some researchers have proposed methods to calibrate multiple RGB cameras that do not have any overlapping regions, which use the reflection of a mirror plane to generate virtual overlap regions [18,19]. These methods can achieve our goal thanks to the correspondences between color images that are obtained from the virtual overlap regions. However, it is necessary to move a mirror plane over again and again to acquire a sufficient number of correspondences because they are derived only from the area that a mirror plane can cover. Considering this effort, the more widely the Kinects are located, the more impractical the methods become. Therefore, we propose a practical method to calibrate extrinsic parameters of such Kinects. In our method, we add extra RGB cameras in an environment to solve the shortage of overlapping regions. In the overlapping regions among Kinects and extra RGB cameras, we can obtain a sufficient number of corresponding points between color images to estimate their camera parameters (Fig. 1(b) top). Compared with the methods using a mirror plane, our method is much more practical because it is unnecessary to move them once they were mounted at their proper location. Although we might obtain 3D corresponding points between range data by using extra range sensors instead of the RGB cameras, we consider that our method is more suitable because the observation area of color images of a Kinect is larger than that of range data, which is accurate within only several meters from a Kinect [15]. Next, because there are independent errors in color images and range data of Kinects, we consider the optimization of the estimated extrinsic parameters by minimizing not only the errors of color images but also those of range data. One well-known optimization approach for range data is the alignment between range data such as iterative closest point [20]. However, due to the accumulated error of each neighboring alignment, the whole range data would not be aligned properly. Therefore, we employ an optimization approach of the alignment between the range data of each Kinect and a 3D world coordinate space. In our method, planar regions, which exist in a general environment such as floors and walls, are used. The position of these regions can be easily measured in world coordinates beforehand. Using range data of the planar regions and their measured position, we minimize the distance between their position and their 3D points projected from range data as well as the error of correspondences in color images (Fig. 1b bottom).

3. Algorithm

This section describes the details of our calibration method. In this section, we continue to regard that two cameras of a Kinect

have same intrinsic and extrinsic parameters, as was stated in Section 2.

3.1. Calibration of intrinsic parameters As mentioned in Section 2, we calibrate both intrinsic camera parameters and parameters to compensate range errors as the intrinsic calibration. To calibrate intrinsic camera parameters, we employ Zhang's method [8] to each Kinect. To compensate range errors, we define an error model as (1) on the basis of the relationship that an error in the IR camera coordinates yields a measured range error that increases quadratically with the distance [15].

$$e(\boldsymbol{m}) = a(\boldsymbol{m})d(\boldsymbol{m})^2 \tag{1}$$

where $a(\boldsymbol{m})$, $d(\boldsymbol{m})$, and $e(\boldsymbol{m})$ are the error parameter, the observed range value, and its error at a pixel $\boldsymbol{m} = (u, v)$, respectively. We consider that each Kinect has its own error parameters because we found through observation that an error distribution of range data of a Kinect differs from those of others. Once the error parameters are obtained, the errors are corrected by

$$D(\boldsymbol{m}) = d(\boldsymbol{m}) - e(\boldsymbol{m}) = d(\boldsymbol{m}) - a(\boldsymbol{m})d(\boldsymbol{m})^2 \tag{2}$$

where $D(\boldsymbol{m})$ denotes the compensated range value at the pixel $\boldsymbol{m}$.

To estimate the error parameters, we apply the following procedure to each Kinect. First, a Kinect is located at various positions with its camera view perpendicular to a planar wall. After capturing its range data at each position, we obtain the parameter of each pixel by the least-squares method.

3.2. Initial estimation of extrinsic parameters We initially estimate extrinsic parameters, that is, the rotation $\boldsymbol{R}$ and translation $\boldsymbol{t}$ of each Kinect. As mentioned in Section 2, in the arrangement of Kinects with consideration of the interference problem, it is impossible to apply existing methods straightforwardly because their overlapping regions are not large enough to supply a sufficient number of correspondences between color images. Therefore, in our method, we add some extra RGB cameras in an environment to obtain a sufficient number of corresponding points between color images. Using the corresponding points, we estimate their extrinsic parameters as follows. First, extrinsic parameters are obtained by the direct linear transform technique. Then, bundle adjustment [9] is performed to improve the extrinsic parameter accuracy.

In the above estimation, it only achieves weak calibration that leaves the scale indeterminacy of the extrinsic parameters because there is no information about a metric scale in the estimation. Thus, we prepare some markers whose position is measured in world coordinates beforehand to determine their metric scale.

3.3. Optimization of extrinsic parameters We optimize the extrinsic parameters that are initially estimated in Section 3.2 by using not only the correspondences between color images but also the range data of planar regions. Since the intrinsic parameters are obtained accurately beforehand, we fix their value so that the number of optimized parameters can be decreased for stable optimization. To acquire the optimal parameters that minimize both sensor errors simultaneously, we minimize the objective function of (3) by the Levenberg–Marquardt algorithm.

$$\hat{\boldsymbol{R}}_1^{(\lambda)}, \hat{\boldsymbol{t}}_1^{(\lambda)}, \cdots, \hat{\boldsymbol{R}}_N^{(\lambda)}, \hat{\boldsymbol{t}}_N^{(\lambda)} = \underset{\boldsymbol{R}_1, \boldsymbol{t}_1, \cdots, \boldsymbol{R}_N, \boldsymbol{t}_N}{\operatorname{argmin}} \left((1-\lambda)E_l + \lambda E_p\right). \tag{3}$$

In (3), E_l is the error about the corresponding points between color images captured from Kinects and extra cameras, and E_p is the error about planar regions of range data of Kinects. These details are mentioned in the next paragraph. λ is the weighting coefficient value of two error values to consider the different effect between E_l and E_p in infinitesimal change of the optimized extrinsic parameters. Because the optimized result varies depending on the value of λ, it is necessary to use an appropriate one. Therefore, in our method, first we prepare some candidates of extrinsic parameters that are obtained by optimization with different values of λ. Then, from the candidates, we adopt appropriate extrinsic parameters when both errors are minimal simultaneously.

E_l is defined as (4) to evaluate the distance between a corresponding point and the epipolar line. The distance is calculated as the absolute value of the inner product of the unit normal vector of the epipolar plane and the displacement vector between two cameras. In (5), it is normalized by the length between two cameras to remove the influence of the length.

$$E_l = \frac{1}{N_{\text{cor}}} \sum_{k=1}^{N_{\text{cor}}} \left(\frac{1}{C_{\text{cam}}^{(k)}} \sum_{i=1, j=i+1}^{N_{\text{cam}}} l_{ij}^{(k)} \right) \tag{4}$$

$$l_{ij}^{(k)} = \frac{\left| \left(\boldsymbol{v}_i^{(k)} \times \boldsymbol{v}_j^{(k)} \right) \cdot \boldsymbol{t}_{ij} \right|}{\left\| \boldsymbol{v}_i^{(k)} \times \boldsymbol{v}_j^{(k)} \right\| \left\| \boldsymbol{t}_{ij} \right\|} \tag{5}$$

where N_{cam} is the total number of cameras, which is equal to the sum of the number of Kinects and extra cameras. N_{cor} is the total number of corresponding points between color images captured from the cameras. $C_{\text{cam}}^{(k)}$ is the number of pairwise combinations from cameras that observe the kth corresponding point. $\boldsymbol{v}_i^{(k)}$ and $\boldsymbol{v}_j^{(k)}$ are the vectors from the focal point of the ith and the jth cameras to the kth corresponding point in their own normalized image plane, respectively. $\boldsymbol{t}_{ij}$ is the displacement vector from the translation $\boldsymbol{t}_i$ to $\boldsymbol{t}_j$.

On the other hand, E_p is calculated from the average of p_j, denoting the range error. It is defined as the average of the distances between the measured position of a planar region Π and its corresponding 3D point projected from range data, expressed as follows:

$$E_p = \frac{1}{N_{\text{kin}}} \sum_{j=1}^{N_{\text{kin}}} \sum_{i=1}^{N_\Pi} \frac{1}{S_{\Pi_i}} p_j\left(\Pi_i\right) \tag{6}$$

$$p_j\left(\Pi_i\right) = \sum_{\boldsymbol{X}_O(\boldsymbol{m}) \in \Pi_i} h\left(\boldsymbol{X}_O\left(\boldsymbol{m}\right), \Pi_i\right) \tag{7}$$

where N_{kin} and N_Π are the number of Kinects and the planar regions, respectively. S_{Π_i} is the area of the ith planar region. $\boldsymbol{X}_O\left(\boldsymbol{m}\right)$ is the 3D world coordinate point that is projected from the range data of the jth Kinect, and h is the metric distance function between a 3D point and its planar position.

4. Experiments

4.1. Experimental setup For experiments, we used an indoor environment that was designed as a corridor (Fig. 2a). In this environment, we mounted some Kinects facing each other across the corridor at 2.1 m height. Their camera views were set downward, as shown in the yz cross-section view of Fig. 2a so that they could observe pedestrians in this corridor. Consequently, the laser source of a Kinect was not included in the views of the others, which meant that we did not have to consider the influence of the appearance of these laser sources.[1] To define a distance between neighboring Kinects that was adequate to capture range data of a pedestrian with few missing regions, we observed

[1] We also confirmed through different observations that even though a laser source is emitted into a Kinect, it does not affect the range data while only range values of the corresponding pixels are missing in the captured range data.

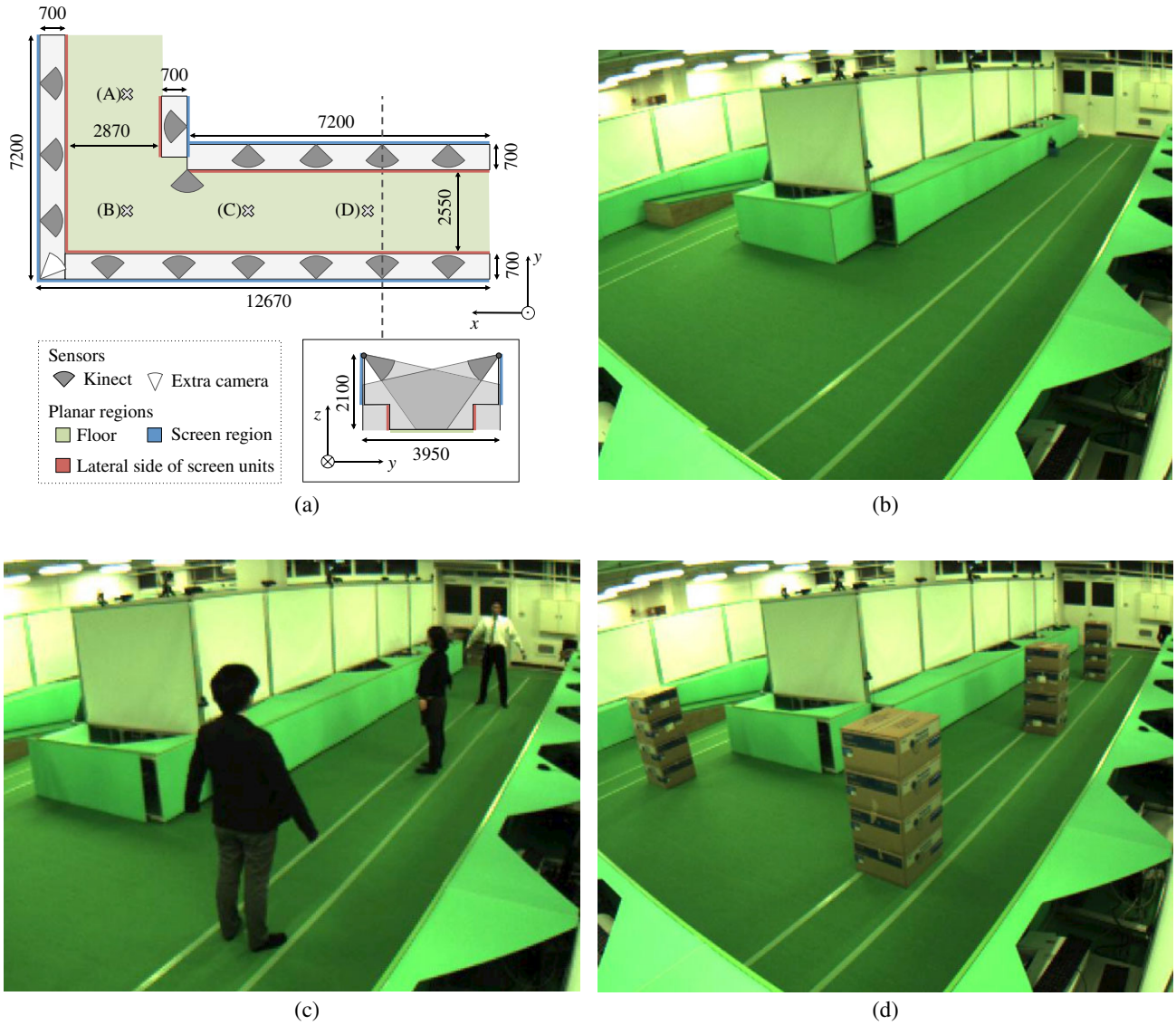


Fig. 2. View of the experimental environment. (a) Layout drawing, (b) Sample image captured from the extra camera, (c) Human bodies, and (d) Cardboard boxes

how the interference appeared on range data of the region of a person who walked in this environment while changing the distance from 3.0 to 0.5 m at 0.5 m intervals. As a result, 2.0 m was the shortest distance that satisfied the condition that there were few missing data on his/her region. Therefore, we employed the distance between neighbors as 2.0 m. To control these Kinects and save captured data, each Kinect was connected to its own computer, whose clock was adjusted by an NTP server.

For the range error calibration described in Section 3.1, we prepared a special setting where a Kinect could be located perpendicular to a wall. In this setting, we captured range data of each Kinect while changing the distance between the Kinect and the wall from 1.0 to 3.0 m at 0.5 m intervals before mounting it in the environment.

For the initial estimation of extrinsic parameters (Section 3.2), one wide-range RGB camera, which was enough to cover the whole environment, was additionally mounted at the left bottom corner of Fig. 2a as an extra camera. An image captured from the extra camera is shown in Fig. 2b. To acquire corresponding points between color images, we first captured images of an LED light in poor illumination and recorded their timestamp. In this capture, we swung the light as widely as possible so that it could be widely distributed in the environment. Then, to reduce the gap of capture timing, we virtually generated corresponding points at specified time intervals from captured ones. In the generation of a corresponding point at a given time, we linearly interpolated the image coordinate values of captured corresponding points by using their timestamp that was recorded in a common time system due to the NTP server. As stated in Section 3.2, to represent calibrated parameters in metric world coordinates, it is necessary to prepare some corresponding points whose position is known in metric world coordinates. In this experiment, we prepared eight corresponding points located at two different heights of four positions indicated by (A)–(D) in Fig. 2a.

For the optimization of extrinsic parameters (Section 3.3), we use three types of planes, that is, the floor, screen regions, and the lateral side of screen units. They are represented by colors in Fig. 2a. To obtain the range data of the planar regions, we first captured one where no foreground objects appeared, which we call background data. In this capture, we could obtain the full range data with no effect of interference because the data were captured from each individual Kinect with it turned on while the others were turned off. Then, we manually clipped the range data of the planar regions from the background data. The metric position of the planar regions was derived from the layout map with a high precision of ± 0.5 cm because the environment

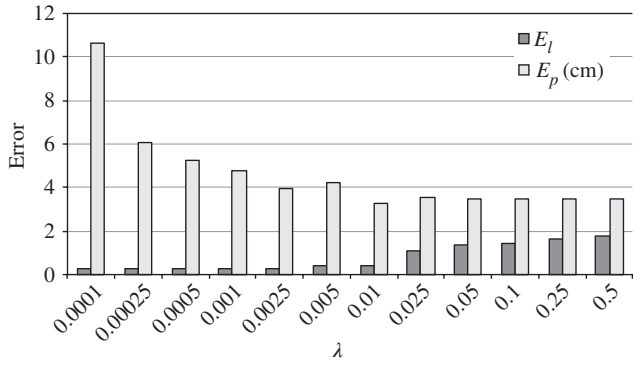


Fig. 3. Relation between λ and errors

was built very precisely by using surveying instruments and laser markers.

4.2. Changing the Weighting Coefficient of Two Errors In the optimization of extrinsic parameters, as explained in Section 3.3, we first prepare some candidates of extrinsic parameters that are obtained by the optimization with different values of λ. Then, we adopt suitable extrinsic parameters of when both errors are minimal simultaneously from the candidates. In this experiment, we prepared some values of λ from 0.5 to 0.0001. Figure 3 shows E_l and E_p of each value of λ. When deciding λ when both errors are minimal simultaneously from the candidates, we used the following method based on the relation that a smaller λ leads to a smaller E_l and, conversely, a larger λ results in a smaller E_p. First, we calculated the change rate of each error between neighboring λ. Then, we adapted λ that minimized the sum of the change ratio of both errors, which was 0.01.[2] The suitable extrinsic parameters of $\lambda = 0.01$ were subsequently used in all the experiments.

4.3. Experimental Results Our extrinsic calibration method comprises the initial estimation of extrinsic parameters (Section 3.2) and the optimization of extrinsic parameters (Section 3.3). In these experiments, we focus on the difference between the results before and after optimizing extrinsic parameters to confirm the effectiveness of the whole proposed method including the optimization. Note that we attempted the initial estimation without using the extra camera, and confirmed that it did not work well due to an insufficient number of corresponding points between color images of only the Kinects; the estimated position of Kinects was obviously incorrect. We also attempted the initial estimation without using the corresponding points between one Kinect and another. As a result, it also failed. From these results, it can be concluded that in these experiments, it was essential for the initial estimation to use both larger overlap regions between the extra camera and the Kinects and smaller ones between the Kinects.

To evaluate our method, we performed shape reconstruction of the environment where people and cardboard boxes were located as foreground objects by registering range data of all the Kinects in world coordinates. In qualitative evaluation, we used reconstruction data of the environment where three persons were at positions (B)–(D) of Fig. 2a, as shown in Fig. 2c. In quantitative comparison, we used reconstruction data of cardboard boxes whose size was carefully measured at the same three positions and the position (A), as shown in Fig. 2d.

Figure 4 shows the reconstruction data of the environment with the persons before and after optimizing extrinsic parameters. As seen from the left image of Fig. 4a, the shape of a target object at the position (B) was far from a human body before optimization. By contrast, after optimization, the shape of the person was well reconstructed. Figure 4b shows their close-up images, whose view is represented as the red frame in Fig. 4a. In the result before optimization, we notice that edges of a side wall, which should be straight, were also incorrectly reconstructed. By contrast, after optimization the edges of the side wall became almost straight. From these results, it can be confirmed that our optimization procedure worked well.

Although the above reconstruction results proved the effectiveness of the optimization, there remain concerns that the bad reconstruction results before optimization were caused by the failure of the initial parameter estimation. To confirm whether the initial estimation procedure had problems, we show the average reprojection error of the corresponding points and the camera viewpoints in Fig. 5a. From this figure, we can see that the camera viewpoints were estimated almost correctly and there were few reproduction errors of all corresponding points. Therefore, it can be concluded that the initial estimation procedure worked well without any problems thanks to a sufficient number of the corresponding points obtained from the Kinects and the extra camera. Totally considering the results of both Fig. 4 and Fig. 5, we can say that it was necessary for Kinect calibration to minimize the error of not only color images but also that of range data even though the error of color images were increased.

Next, we show the result of a quantitative evaluation using the cardboard boxes located at the positions (A)–(D) as shown in Fig. 2d. Before shape reconstruction including these boxes, we first measured their size with a ruler. Next, we marked the corner positions where a box should be located on the ground by using a measure and a leveling string. Then, we carefully located them so that their corners were aligned with their marked positions. The measuring precision was estimated at about ±1.0 cm. After the shape reconstruction, we manually extracted the 3D point data of each surface of each box from the reconstructed data. The extract box shape consisted of about 60 000 point data. For quantitative evaluation, we calculated the distance between each extracted surface data and its corresponding planner position as the error.

Figure 6 represents the average error of each box. From the result, it is obvious that the parameter optimization successfully improved the reconstruction accuracy of all boxes including the one at the position (A), which was difficult to achieve with high accuracy because it was placed at the edge of a scene. In position (B), its errors were dramatically decreased. In Fig. 2a, it is found that cameras did not surround this position. After the parameter optimization, we could recover poor estimation derived from these badly located cameras. In position (C), the accuracy was the best of 2.5 cm after the parameter optimization. It is because the extra camera located near there could supply reliable corresponding points and because Kinects surrounded there. It was previously reported that when the distance from a Kinect to the subject is 3 m, the random error and the range resolution are about 1.5 and 2.5 cm on average, respectively [15]. Considering the total error caused by the above two factors, it can be said that the reconstruction error was almost zero.

4.4. Total time required for calibration When we used a commercial computer that had Intel® Core™ i7 processors (8 cores, 2.93 GHz) and 16 GB of memory for the calibration, it took nearly 4 h to complete the whole procedures. We spent

[2] λ might be included in parameters that should be optimized. However, when we tried to optimize the value of λ together with the extrinsic parameters, we obtained only unreasonable parameters with quite a small value of λ. This is because E_l was intensively minimized and E_p was ignored by the small value of λ to minimize the objective function.

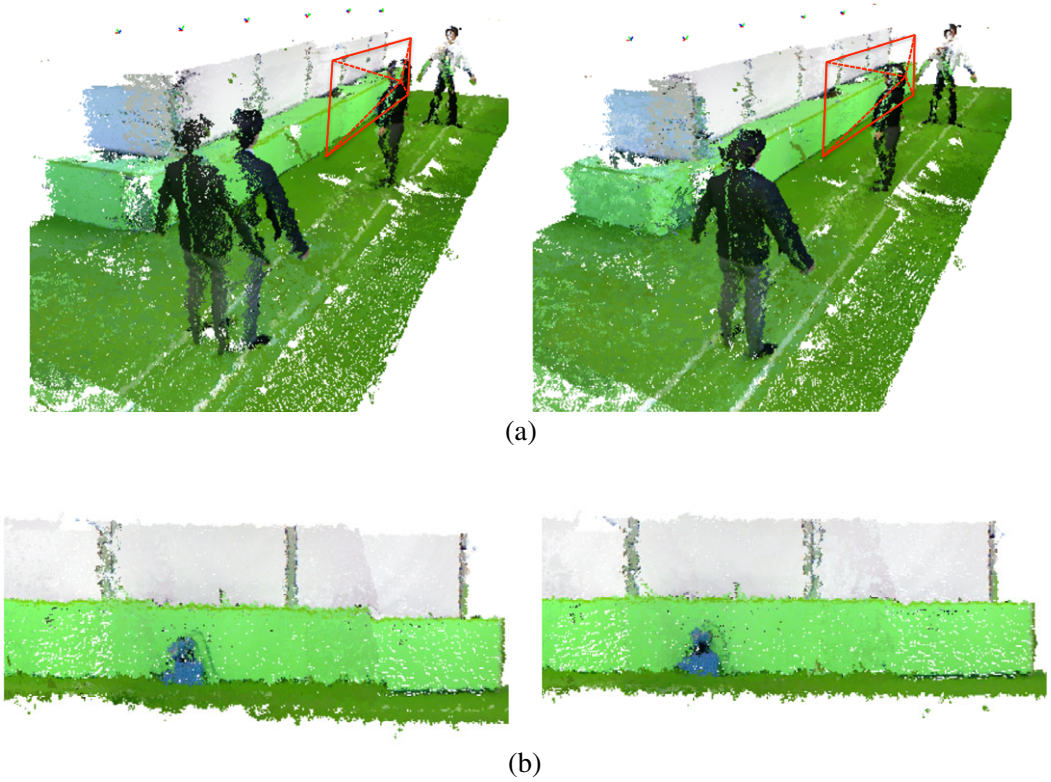
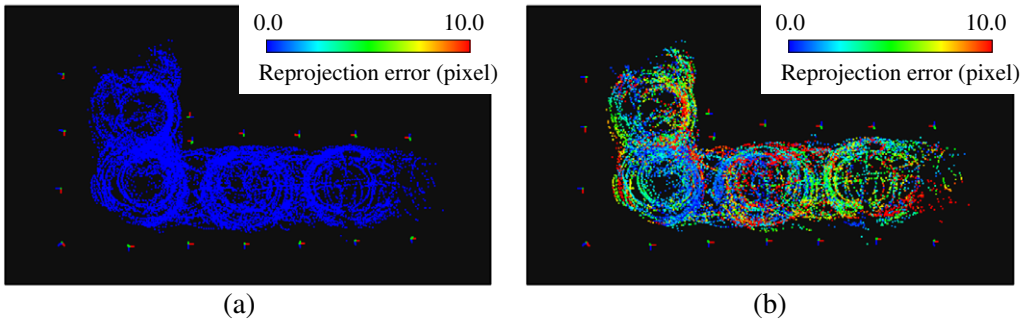


Fig. 4. Reconstructed shapes of the experimental environment (Left and right: Before and after extrinsic parameter optimization). (a) Overview of reconstructed shapes of the experimental environment. The red frame represents the view of Fig. 4b and (b) Close-up view of reconstructed shapes of the experimental environment



(a) (b)

Fig. 5. Average reprojection error of corresponding points and camera viewpoints. (a) Before optimizing extrinsic parameters and (b) After optimizing extrinsic parameters

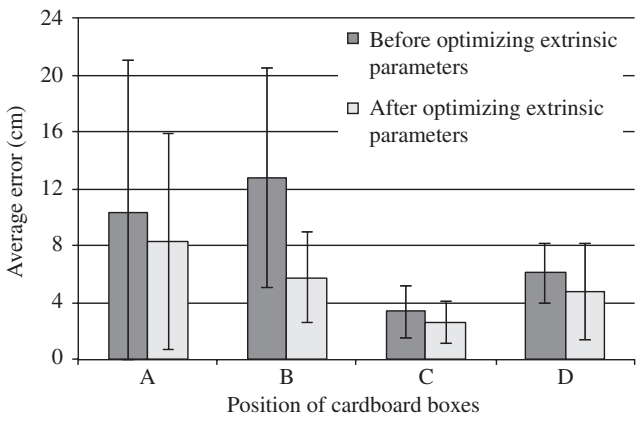


Fig. 6. Average error between the measured position of box surfaces and their 3D points projected from range data. The standard deviation of errors is depicted by the line on the bars

most of the time to extract the range data of the planar regions manually and to complete the optimization of extrinsic parameters while changing the value of λ. If we can extract the range data of planar regions automatically and implement parallel processing to find optimal λ, the time will be shortened.

5. Conclusion

In this paper, we proposed an accurate and efficient calibration method for multiple Kinects whose overlapping regions are not large enough. For their extrinsic calibration, we first added extra RGB cameras in an environment to solve the shortage of overlapping regions. Thanks to the overlapping regions among Kinects and extra RGB cameras, we could obtain a sufficient number of corresponding points between color images to estimate their camera parameters. Then, the estimated parameters were optimized to minimize both the errors of corresponding points between color images and the ones of range data of planer regions, which existed in a general environment. This procedure was

suitably designed based on the characteristics of the color images and range data. From the results of the experiments, we confirmed the effectiveness of our method.

As future work, we will try to reduce the time required for calibration by achieving the improvement described in Section 4.4. Moreover, we will try to generate synchronous range data. Because there is no function to capture range data synchronously in Kinects, a temporal capture gap occurs among them. To deal with this issue, we will try to generate range data at a common time of Kinects by a morphing technique for 3D range data [21].

Acknowledgment

This work was partly supported by the JST CREST "Behavior Understanding based on Intention-Gait Model" project.

References

(1) http://www.xbox.com/en-US/kinect. Accessed on July 16, 2014.
(2) Spinello L, Arras KO. People Detection in RGB-D Data. Proceedings of 2011 IEEE/RSJ International Conference on the Intelligent Robots and Systems (IROS2011), 2011; 3838–3843.
(3) Nakajima H, Mitsugami I, Yagi Y. Depth-based gait feature representation. *IPSJ Transactions on Computer Vision and Applications* 2013; **5**:94–98.
(4) Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, Kipman A, Blake A. Real-time human pose recognition in parts from single depth images. Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011), 2011; 1297–1304.
(5) Berger K, Ruhl K, Brümmer C, Schröder Y, Scholz A, Magnor M. Markerless motion capture using multiple color-depth sensors. Proceedings of the 16th International Workshop on Vision, Modeling and Visualization (VMV2011), 2011; 317–324.
(6) Izadi S, Newcombe RA, Kim D, Hilliges O, Molyneaux D, Hodges S, Kohli P, Shotton J, Davison AJ, Fitzgibbon A. KinectFusion: real-time dynamic 3D surface reconstruction and interaction. Proceedings of SIGGRAPH Talks, 2011; 23pp.
(7) Weiss A, Hirshberg D, Black M. Home 3D body scans from noisy image and range data. Proceedings of the 13th International Conference on Computer Vision (ICCV2011), 2011; 1951–1958.
(8) Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions Pattern Analysis and Machine Intelligence (TPAMI)* 2000; **22**(**11**):1330–1334.
(9) Triggs B, McLauchlan P, Hartley R, Fitzgibbon A. Bundle adjustment - a modern synthesis. *Vision Algorithms: Theory and Practice* 2000; **1883**:298–372.
(10) Auvinet E, Meunier J, Multon F. Multiple depth cameras calibration and body volume reconstruction for gait analysis. Proceedings of the 11th International Conference on the Information Science, Signal Processing and their Applications (ISSPA2012), 2012;
478–483.
(11) Macknojia R, Chávez-Aragón A, Payeur P, Laganiáre R. Calibration of a network of Kinect sensors for robotic inspection over a large workspace. Proceedings of the IEEE Workshop on Robot Vision (WORV), 2013; 184–190.
(12) Kim JH Choi JS, Koo BK. Calibration of multi-Kinect and multi-camera setup for full 3D reconstruction. Proceedings of the 44th International Symposium on Robotics (ISR2013), 2013; 1–5.
(13) Staranowicz A, Brown GR, Morbidi F, Mariottini GL. Automatic extrinsic calibration of RGB-D system using two views of natural scene. Proceedings of the 6th Pacific-Rim Symposium on Image and Video Technology (PSIVT), 2013; 265–278.
(14) Devaux JC, Abdelkader HH, Colle E. Fully automatic extrinsic calibration of RGB-D system using two views of natural scene. Proceedings of the 13th International Conference on Control Automation Robotics and Vision (ICARCV2014), 2014; 894–900.
(15) Khoshelham K, Elberink SO. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* 2012; **12**(**2**): 1437–1454.
(16) Klette R, Schluns K, koschan A. *Computer Vision: Three-dimensional Data from Image*. Springer: Singapore; 1998.
(17) Sabata B, Aggarwal JK. Estimation of motion from a pair of range images: a review. *Computer Vision, Graphics, and Image Processing* 1991; **54**(**3**):309–324.
(18) Kumar RK, Ilie A, Frahm JM, Pollefeys M. Simple calibration of non-overlapping cameras with a mirror. Proceedings of the 21st IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008), 2008; 1–7.
(19) Lébraly P, Deymier C, Aider OA, Royer E, Dhome M. Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: application to vision-based robotics. Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), 2010; 5640–5647.
(20) Besl PJ, McKay ND. A method for registration of 3-D shapes. *IEEE Transactions Pattern Analysis and Machine Intelligence (TPAMI)* 1992; **14**(**2**):239–256.
(21) Nakajima H, Makihara Y, Hsu H, Mitsugami I, Nakazawa M, Yamazoe H, Habe H, Yagi Y. Point cloud transport. Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, 3803–3806.

Mitsuru Nakazawa (Non-member) received the B.E. and M.E. degrees from Shibaura Institute of Technology in 2006 and 2008, respectively, and the Ph.D. degree in Engineering from Keio University, in 2011. From 2008 to 2011, he was a JSPS Research Fellow for Young Scientists (DC1). From 2011 to 2015, he was a post-doctoral researcher at the Institute of Science and Industrial Research, Osaka University. He is currently working as a scientist with Rakuten Institute of Technology, Rakuten, Inc. His research interests include sensing techniques for object measurement and data visualization based on computer vision and pattern recognition. Dr Nakazawa is a member of the IEEE and IPSJ.

Ikuhisa Mitsugami (Non-member) received the B.S. degree in Engineering from Kyoto University in 2001, and the M.S. and Ph.D. degrees in Engineering from Nara Institute of Science and Technology in 2003 and 2007, respectively. He is currently an Assistant Professor with the Institute of Scientific and Industrial Research, Osaka University. His research interests include geometry in computer vision, detection and tracking, gait analysis. Dr Mitsugami is a member of the IEEE, IEICE, IPSJ, RSJ, and VRSJ.

Hitoshi Habe (Non-member) received the B.E. and M.E. degrees in Electrical Engineering and the D. Info. degree in Intelligence Science and Technology from Kyoto University, Japan, in 1997, 1999, and 2006, respectively. He worked with Mitsubishi Electric Corporation from 1999 to 2002, Kyoto University from 2002 to 2006, Nara Institute of Science and Technology from 2006 to 2011, and Osaka University from 2011 to 2012. He is now a Lecturer with the Department of Informatics, Faculty of Science and Technology, Kinki University, Japan. From 2010 to 2011, he was a Visiting Researcher with the Department of Engineering, University of Cambridge, UK. His research interests include computer vision, pattern recognition, and image processing. Dr Habe is a member of the IEEE, ACM, IEICE, IPSJ, and JSAI.

Hirotake Yamazoe (Non-member) received the B.E., M.E., and Ph.D. degrees in Engineering from Osaka University in 2000, 2002, and 2005, respectively. He was with the Advanced Telecommunications Research Institute International (ATR) during 2005–2011, the Institute of Scientific and Industrial Research, Osaka University, during 2011–2012, and Osaka School of International Public Policy, Osaka University, during 2012–2015. He is now a Lecturer with the College of Information Science and Engineering, Ritsumeikan University. His research interests include computer vision and wearable computing. Dr Yamazoe is a member of the IEICE, IPSJ, ITE, HIS, and ACM.

Yasushi Yagi (Non-member) received the Ph.D. degree from Osaka University in 1991. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a Research Associate in 1990, a Lecturer in 1993, an Associate Professor in 1996, and a Professor in 2003 at Osaka University. He is currently an Executive Vice President, Osaka University. He has served as Chair for many international conferences, including FG1998 (Financial Chair), OMINVIS2003 (Organizing chair), ROBIO2006 (Program co-chair), ACCV2007 (Program chair), PSVIT2009 (Financial chair), ICRA2009 (Technical Visit Chair), ACCV2009 (General chair), ACPR2011 (Program co-chair) and ACPR2013 (General chair). He also served as the Editor of IEEE ICRA Conference Editorial Board (2007–2011). He is an Editorial member of IJCV and the Editor-in-Chief of the IPSJ Transactions on Computer Vision & Applications. He was awarded ACM VRST2003 Honorable Mention Award, IEEE ROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 Finalist for Best Vision Paper, MIRU2008 Nagao Award, and PSIVT2010 Best Paper Award. His research interests include computer vision, medical engineering, and robotics. Prof. Yagi is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.