Extrinsic Camera Calibration Without Visible Corresponding Points Using Omnidirectional Cameras

Shogo Miyata, *Student Member, IEEE*, Hideo Saito, *Senior Member, IEEE*, Kosuke Takahashi, Dan Mikami, *Member, IEEE*, Mariko Isogawa, and Akira Kojima

***Abstract*— This paper proposes a novel algorithm that calibrates multiple cameras scattered across a broad area. The key idea of the proposed method is "using the position of an omnidirectional camera as a reference point." The common approach to calibrating multiple cameras assumes that the cameras capture at least some common points. This means calibration becomes quite difficult if there are no shared points in each camera's field of view (FOV). The proposed method uses the position of an omnidirectional camera to determine point correspondence. The position of an omnidirectional camera relative to the calibrated camera is estimated by the theory of epipolar geometry, even if the omnidirectional camera is placed outside the camera's FOV. This property makes our method applicable to multiple cameras scattered across a broad area. Qualitative and quantitative evaluations using synthesized and real data, e.g., a sports field, demonstrate the advantages of the proposed method.**

***Index Terms*— Camera calibration, omnidirectional camera, non-overlapping cameras.**

I. Introduction

MULTI-CAMERA calibration, which is commonly realized by using the corresponding points observed by several cameras simultaneously, is one of the most fundamental techniques in computer vision [1], [2] and is necessary for various applications including car-mounted cameras [3], robot control [4], [5], AR [6] and free-viewpoint video [7]–[9]. In our study, we mainly focus on the extrinsic calibration of a set of cameras that cover a broad area, e.g., surveillance cameras deployed in an urban area and camera set in a sports field.

When several cameras are deployed over a wide field, quite often their FOVs do not overlap. In addition, they sometimes do not have corresponding points, even when their FOVs overlap.

The second difficulty is scaling. When we calibrate cameras using a calibration or reference board, it should be captured larger than a certain size. Thus, to calibrate cameras that cover wide areas, the calibration board should be so large as to be impractical.

To solve the first problem, *i.e.*, cameras' FOVs are not shared, Takahashi *et al.* [10] and Rodrigues *et al.* [11] use planar mirrors. First, they generate views shared by the cameras through the reflections from the mirror. Then, they place reference objects on the shared view. As another solution, Caspi and Irani [12] and Esquivel *et al.* [13] use cameras mounted on an assembly jig and solve structure-from-motion under the constraint that the cameras on the jig exhibit the same relative position and rotation. However, because these solutions require additional devices, they cannot be applied to cameras scattered across wide areas.

To solve the latter problem, *i.e.*, cameras scattered in a large field, Kitahara *et al.* [14] propose a method that uses 3D laser-surveying instruments. They estimate the relative position and rotation of cameras from a measured 3D model of the field. However, 3D laser-surveying instruments are too expensive for casual use. In some cases, capturing the 3D model before calibration is impossible.

Our study addresses these problems by using an omnidirectional camera. The key idea of our method is "using the position of an omnidirectional camera as the reference point." First, the proposed method estimates the projection point of the omnidirectional camera on the basis of the essential matrix linking each camera to the omnidirectional camera. Next, it estimates extrinsic parameters using the projection points as the corresponding points between cameras. Note that the omnidirectional camera does not need to be observed by each camera. Even if it is placed outside of the camera's FOV, its projection point can be obtained by epipolar geometry.

This method has four advantages: (1) because the position of the omnidirectional camera can be used as a reference point, it does not require the corresponding point to be within any shared view; (2) because it only requires an omnidirectional camera as the additional device for calibration, it is robust against changes in scale and can be applied to wide-spread cameras; (3) because the extrinsic parameters of each camera

Manuscript received June 14, 2016; revised February 28, 2017, May 27, 2017, and July 3, 2017; accepted July 10, 2017. Date of publication July 25, 2017; date of current version September 13, 2018. This paper was recommended by Associate Editor Y. Wang. *(Corresponding author: Shogo Miyata.)*

S. Miyata and H. Saito are with the Department of Information and Computer Science, Keio University, Yokohama 223-8522, Japan (e-mail: miyata@hvrl.ics.keio.ac.jp; saito@hvrl.ics.keio.ac.jp).

K. Takahashi, D. Mikami, M. Isogawa, and A. Kojima are with NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Yokosuka 239-0847, Japan (e-mail: takahashi.kosuke@lab.ntt.co.jp; mikami.dan@lab.ntt.co.jp; isogawa.mariko@lab.ntt.co.jp; kojima.akira@lab.ntt.co.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2017.2731792

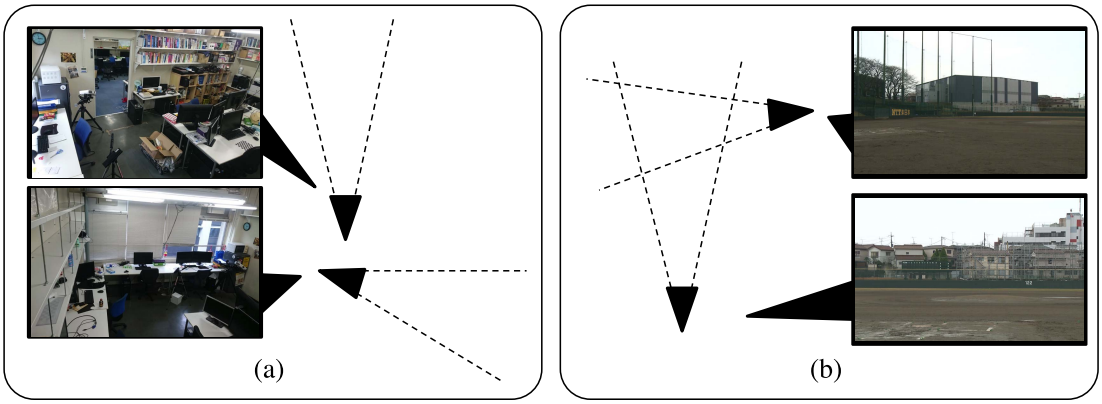

Fig. 1. The target configuration of our method: (a)The case where the camera set does not have a shared view area, (b)The case where the camera set does not have reference points in a shared view area.

can be estimated by the position of epipole of omnidirectional cameras for each camera, we do not need to find any point corresponding to all cameras, but only need to find pairwise corresponding points between each camera and each omnidirectional camera; and (4) due to the rapid popularization of omnidirectional cameras, such as THETA, it is cost effective.

The remainder of this paper is organized as follows. Section II reviews existing calibration methods for cameras without shared views and for cameras widely spread across broad areas and the positions the proposed method among them. Section III proposes the key idea and introduces algorithms that implement the idea. Section IV details evaluations conducted on synthetic data and real data to demonstrate the performance of our method. Section V details the proposed method and Section VI concludes this paper.

II. Related Work

This section reviews existing works from two points of view. First, we introduce multi-camera calibration methods that dispense with corresponding points.

One popular solution is to use planar or spherical mirrors. Planar mirror-based methods [5], [10], [11], [15] set the mirror to make the reference object, which originally lies out of the cameras' view, visible. By using multiple mirror settings, extrinsic camera calibration becomes possible. Spherical mirror-based methods have been proposed [16]. [17] employes the eye-ball as the spherical mirror for calibration. However, our target situation, *i.e.*, calibrating cameras placed across a broad area such as a sports field, poses severe challenges to existing solutions.

Another popular solution involves the motion coherence of cameras mounted on a jig, *i.e.*, camera motions are the same and the relative position and rotation of the cameras are constant. For example, structure-from-motion (SfM) is carried out on the basis of videos captured by cameras mounted on the jig, and then, the restraint condition of the mounted cameras is used for extrinsic calibration [12], [13], [18], [19]. In [20] and [21], SLAM, simultaneous localization and mapping, is used instead of SfM. In a similar manner, [3] uses a pattern placed on a wall and [22] uses a rigid 3D calibration target. All of these studies require a rigid camera jig to make the position and rotation of mounted cameras fixed, which renders them unsuitable for our target situation.

In addition to the above solutions, [23] uses pedestrian trajectories for calibrating surveillance cameras. This method assumes that cameras is placed on the same level, which is difficult to achieve outside. Laser pointers are used for car-mounted cameras that do not share views [24]. References [25] and [26] propose a 1D target. Although many calibration methods have been proposed, they cannot be used for wide coverage of large outside areas.

Second, we introduce existing calibration methods for cameras scattered across broad areas. A calibration board and a 3D laser-surveying instrument are used [14]. This method needs expensive devices and is not good for casual use in terms of cost and rapid deployment. In [27], a 3D model of the play coat is prepared and cameras are calibrated by using the court lines found in the captured image. However, this technique can be applied only to cameras that face downward.

This paper proposes a novel approach that uses an omnidirectional camera for calibrating cameras. The proposed method uses the positions of the omnidirectional camera as the reference points, which are obtained by epipolar geometry between the cameras and the omnidirectional camera. Because the proposed method requires neither mirror nor jig, it effectively supports calibration to cover broad areas.

III. Proposed Method

This section introduces a novel multi-camera calibration method that uses an omnidirectional camera under the situation where there are no reference points in the shared FOV.

A. Problem Definition and Measurement Model

The extrinsic calibration of multiple cameras scattered across a wide area, calibrating such camera is often problematic as their FOVs may not overlap, as illustrated in Fig. 1(a) or there is no 3D reference points for point correspondence in the area shared by each camera's FOV, as illustrated in Fig. 1 (b). We focus on extrinsic camera calibration in such situations. In order to deal with this

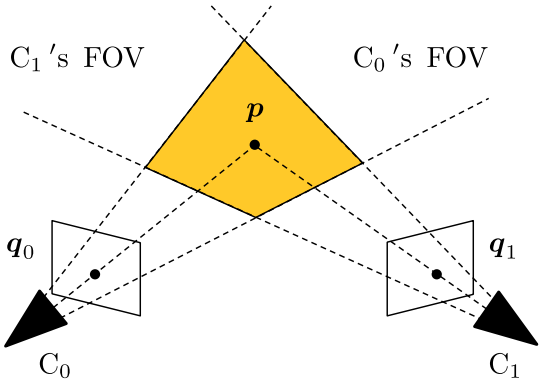

Fig. 2. Epipolar geometry.

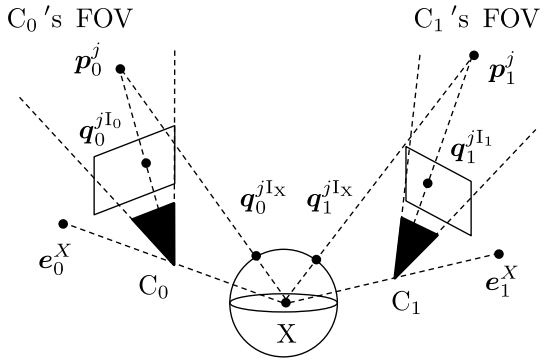

Fig. 3. Measurement model with an omnidirectional camera.

problem, we assume the basic configuration of two cameras $\boldsymbol{C}_0$ and $\boldsymbol{C}_1$ (See Fig. 2).

Assuming that 3D point $\boldsymbol{p}$ is captured by both C_0 and C_1, $\boldsymbol{p}$ is expressed, in each camera's coordinate system, as $\boldsymbol{p}^{\{\mathrm{C}_0\}}$ and $\boldsymbol{p}^{\{\mathrm{C}_1\}}$, respectively. They satisfy the following equation,

$$\boldsymbol{p}^{\{\mathrm{C}_0\}\top}\boldsymbol{E}\,\boldsymbol{p}^{\{\mathrm{C}_1\}} = 0, \tag{1}$$

where $\boldsymbol{E}$ is an essential matrix between C_0 and C_1. $\boldsymbol{p}$ is projected onto image plane $\mathrm{I}_0, \mathrm{I}_1$ as $\boldsymbol{q}_0, \boldsymbol{q}_1$. Given the intrinsic parameters of each camera $\boldsymbol{K}_0, \boldsymbol{K}_1$ and extrinsic parameters $\boldsymbol{R}, \boldsymbol{t}$, which are the camera C_0's poses relative to camera C_1, (1) is expressed as,

$$\boldsymbol{q}_0^{\top}\boldsymbol{K}_0^{-\top}[\boldsymbol{t}]_{\times}\boldsymbol{R}\boldsymbol{K}_1^{-1}\boldsymbol{q}_1 = 0, \tag{2}$$

where $[\boldsymbol{t}]_{\times}$ means the skew-symmetric matrix of $\boldsymbol{t}$. The goal of this research is to estimate extrinsic parameters $\boldsymbol{R}$ and $\boldsymbol{t}$.

The general approach to estimating these extrinsic parameters is to use point correspondence such as SfM; first, estimate essential matrix $\boldsymbol{E}$ by applying an 8-point algorithm to more than eight pairs of corresponding projections $(\boldsymbol{q}_0, \boldsymbol{q}_1)$, and then decompose $\boldsymbol{E}$ to $\boldsymbol{R}$ and $\boldsymbol{t}$ by singular value decomposition. However, if the cameras to be calibrated are spread over a wide area problems with point correspondence determination arise, since the cameras have no shared FOV, as shown in Fig. 1 (a), and no shared 3D points exist, as shown in Fig. 1 (b). Against this problem, we propose a novel method that uses an omnidirectional camera for extrinsic parameter estimation.

B. Key Idea: The Position of an Omnidirectional Camera as a Reference Point

The key idea of our method is to use the position of an omnidirectional camera as a reference point for the cameras not sharing their FOV.

Our proposed method uses omnidirectional camera X as an additional device, as in Fig. 3. It is expected that there are several 3D points for point correspondence in its shared FOV. Here we denote the N_i 3D points located in the area shared with X and C_i as $\boldsymbol{p}_i^j$ $(i = 0, 1, j = 0, \cdots, N_i - 1)$. They are, in the camera coordinate system and omnidirectional camera coordinate system, expressed as $\boldsymbol{p}_i^{j\{\mathrm{C}_i\}}$ and $\boldsymbol{p}_i^{j\{\mathrm{X}\}}$, respectively. 3D points $\boldsymbol{p}_i^{j\{\mathrm{C}_i\}}$ can be expressed using projection $\boldsymbol{q}_i^{j\{\mathrm{I}_i\}}$, as follows,

$$\boldsymbol{p}_i^{j\{\mathrm{C}_i\}} = s\boldsymbol{K}_i^{-1}\boldsymbol{q}_i^{j\{\mathrm{I}_i\}}, \tag{3}$$

where s is a scale parameter. $\boldsymbol{p}_i^{j\{\mathrm{X}\}}$ can be expressed using equi-rectangular projection $\boldsymbol{q}_i^{j\{\mathrm{X}\}} = (u_i^j, v_i^j)^{\top}$, as follows,

$$\boldsymbol{p}_i^{j\{\mathrm{X}\}} = f(\boldsymbol{q}_i^{j\{\mathrm{I_X}\}}) = (\sin(\phi)\cos(\theta), \cos(\phi), \sin(\phi)\sin(\theta))^{\top}, \tag{4}$$

where θ and ϕ are the angles as $\theta = \frac{2\pi}{\mathrm{W}}(u_i^j - \frac{\mathrm{W}}{2})$, $\phi = \frac{\pi}{\mathrm{H}}v_i^j$, where W and H are the width and height of the omnidirectional image respectively. These parameters satisfy following equation,

$$\boldsymbol{q}_i^{j\{\mathrm{I}_i\}\top}\boldsymbol{K}_i^{-\top}\boldsymbol{E}_i^{\mathrm{X}}f(\boldsymbol{q}_i^{j\{\mathrm{I_X}\}}) = 0, \tag{5}$$

where $\boldsymbol{E}_i^{\mathrm{X}}$ is an essential matrix between C_i and X, and it can be computed by applying an 8-point algorithm to more than eight pairs of $\boldsymbol{q}_i^{j\{\mathrm{I}_i\}}$ and $\boldsymbol{q}_i^{j\{\mathrm{I_X}\}}$.

The translation vector of omnidirectional camera $\boldsymbol{t}^{\mathrm{X}}$ can be computed by applying singular value decomposition to $\boldsymbol{E}_i^{\mathrm{X}}$, and this $\boldsymbol{t}^{\mathrm{X}}$ is projected onto C_i's image plane I_i as epipole $\boldsymbol{e}_i^{\mathrm{X}}$ (Fig. 3). Even if $\boldsymbol{e}_i^{\mathrm{X}}$ is out of C_i's field of view, it can be computed as a point on the virtual image plane.

Omnidirectional camera X has a wide FOV and so shares the FOVs of all cameras C_i, so epipole $\boldsymbol{e}_i^{\mathrm{X}}$ on the image planes of all cameras can be computed. This means that the position of omnidirectional camera X itself becomes the shared reference point for all cameras C_i.

C. Estimation of Extrinsic Parameters

In this section, we detail an extrinsic calibration algorithm that implements our key idea. Here the inputs for our proposed algorithm are the images captured by cameras C_0 and C_1, that is I_0 and I_1, and images captured by omnidirectional camera X, that is $\mathrm{I}_{\mathrm{X}_k}(k = 0, \cdots, N_{\mathrm{X}} - 1)$. Notice that $\mathrm{I}_{\mathrm{X}_k}$ is captured at different positions.

1) Algorithm1: This algorithm consists of two steps: (Step1) obtain the pair of corresponding points in I_0 and I_1, (Step2) estimate extrinsic parameters $\boldsymbol{R}$ and $\boldsymbol{T}$ from the pair of correspondences.

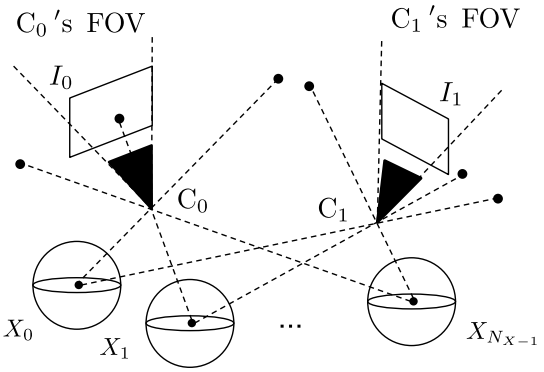

Fig. 4. Step1 of Algorithm1.

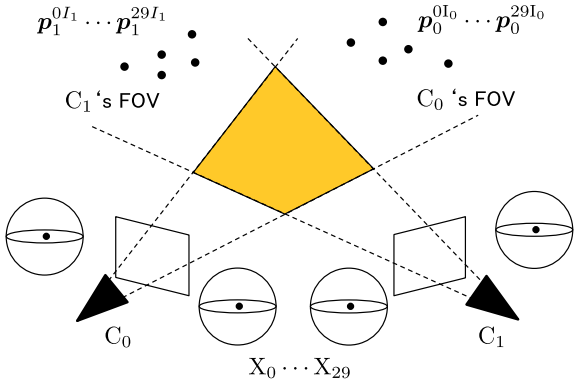

Fig. 5. Configuration for evaluation using synthetic data.

a) (Step1): First, we obtain the pairs of corresponding points in I_0 and $\mathrm{I}_{\mathrm{X}_k}$ and compute essential matrix $\boldsymbol{E}_0^{\mathrm{X}_k}$ by applying an 8-point algorithm to them. Second, we compute translation vector $\boldsymbol{t}_0^{\mathrm{X}_k}$ by calculating the singular value decomposition of $\boldsymbol{E}_0^{\mathrm{X}_k}$ and obtain epipole $\boldsymbol{e}_0^{\mathrm{X}_k}$ by projecting $\boldsymbol{t}_0^{\mathrm{X}_k}$ onto the image plane. Next, we apply same process to I_1 and $\mathrm{I}_{\mathrm{X}_k}$ and obtain $\boldsymbol{e}_1^{\mathrm{X}_k}$. Here, $\boldsymbol{e}_1^{\mathrm{X}_k}$ is the point that corresponds to $\boldsymbol{e}_0^{\mathrm{X}_k}$.

By applying the above process to all $\mathrm{I}_{\mathrm{X}_k}$, we obtain N_{X} pairs of corresponding points between I_0 and I_1, as shown in Fig. 4.

b) (Step2): We compute essential matrix $\boldsymbol{E}_0^1$ between C_0 and C_1 by applying an 8-point algorithm to N_{X} pairs of corresponding points obtained in (Step1). Finally, we calculate the singular value decomposition of $\boldsymbol{E}_0^1$ and obtain extrinsic parameters $\boldsymbol{R}$ and $\boldsymbol{t}$ [1].

In this algorithm, N_{X} depends on the required number of pairs of corresponding points for computing $\boldsymbol{E}_0^1$ in (Step2), so in the case of using a 5-point algorithm, N_{X} is more than or equal to five.

This algorithm well estimates the correct extrinsic parameters if the environment is noiseless. As shown in Fig. 6, however, extrinsic parameter precision degrades remarkably if the input data include observation noise. This is due to the essential matrix estimation step because it has been reported that the essential matrix computation is weak against observation noise [28]; we compute the essential matrix from noisy input data in (Step1) and use this result for computing the other essential matrix in (Step2).

2) Algorithm2: In addition to the key idea proposed in Section III-B, this algorithm introduces the idea of reducing the number of point correspondences used to compute the essential matrix based on [28]. This algorithm estimates the extrinsic parameters of each camera C_i $(i = 0, 1)$ and omnidirectional camera X_k $(k = 0, \cdots, N_{\mathrm{X}} - 1)$ in (Step1) of Algorithm1, without computing the essential matrix (Step2).

a) (Step1): First, we use an 8-point algorithm to estimate essential matrix $\boldsymbol{E}_i^{\mathrm{X}_k}$ between C_i and X_k. Second, we obtain extrinsic parameters of C_i and X_k from $\boldsymbol{E}_i^{\mathrm{X}_k}$ by calculating the singular value decomposition as in [1].

b) (Step2): Using rotation matrices $\boldsymbol{R}_0^{\mathrm{X}_k}$ and $\boldsymbol{R}_1^{\mathrm{X}_k}$, we compute the rotation matrix $\boldsymbol{R}$ between C_0 and C_1 as follows,

$$\boldsymbol{R} = \boldsymbol{R}_1^{\mathrm{X}_k} \boldsymbol{R}_0^{\mathrm{X}_k \top}. \tag{6}$$

TABLE I
COMPARISON OF PROPOSED METHOD

Method	(Step1)	(Step2)
(a)	solve (5) by 8-point algorithm	Algorithm 1
(b)	solve (5) by 5-point algorithm	Algorithm 1
(c)	solve (5) by 8-point algorithm	Algorithm 2
(d)	solve (5) by 5-point algorithm	Algorithm 2

If N_{X} omnidirectional images are used as input, that is, we obtain N_{X} types of $\boldsymbol{R}$, we use the average of them as the output. Since translation vector $\boldsymbol{t}$ has two degrees of freedom, the required N_{X} is two for computing $\boldsymbol{t}$. When $N_{\mathrm{X}} = 2$, $\boldsymbol{t}$ can be computed as follows [28],

$$\boldsymbol{t} = (\boldsymbol{R}\boldsymbol{t}_0^{\mathrm{X}_0} \times \boldsymbol{t}_0^{\mathrm{X}_1}) \times (\boldsymbol{R}\boldsymbol{t}_1^{\mathrm{X}_0} \times \boldsymbol{t}_1^{\mathrm{X}_1}). \tag{7}$$

In case of $N_{\mathrm{X}} > 2$, we compute $\boldsymbol{t}$ as follows,

$$\boldsymbol{t} = \arg\min_{\boldsymbol{t}} \ \Sigma_{k=0}^{N_{\mathrm{X}}-1} (\boldsymbol{R}\boldsymbol{t}_0^{\mathrm{X}_k} \times \boldsymbol{t}_1^{\mathrm{X}_k})^\top \boldsymbol{t}. \tag{8}$$

IV. EXPERIMENTS

This section details the experiments conducted on synthetic and real data sets to evaluate the quantitative and qualitative performance of our method. In the following, we evaluate four variants of the proposed method, as detailed in Table I.

A. Synthesized Data

1) Experiment Environment: We evaluate the impact of observation noise on the performance of the proposed method. Here we use two cameras $\mathrm{C}_0, \mathrm{C}_1$ and one omnidirectional camera X and set them as in Fig. 5. The cameras' intrinsic parameters $\boldsymbol{K}_0$, $\boldsymbol{K}_1$ are as follows,

$$\boldsymbol{K}_0 = \boldsymbol{K}_1 = \begin{bmatrix} 2196.61 & 0 & 799.5 \\ 0 & 2237.36 & 599.5 \\ 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

We set 30 reference points $\boldsymbol{p}_i^j$ $(j = 0, \cdots, 29)$ in the area shared X and each C_i. We capture 30 omnidirectional images $\mathrm{I}_{\mathrm{X}_k} (k = 0, \cdots, 29)$ with a resolution of 5000×2500 pixels. In this evaluation, we add zero-mean Gaussian noise,

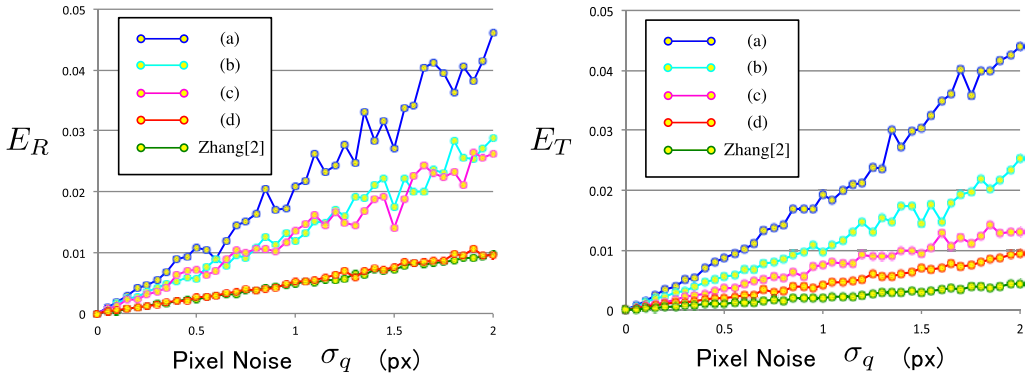

Fig. 6. Estimation error of proposed method under Gaussian noise for $\boldsymbol{q}_i^{j\{\mathrm{I}_i\}}$ and $\boldsymbol{q}_i^{j\{\mathrm{X}_k\}}$ with standard deviation $\sigma_{\boldsymbol{q}}$.

whose standard deviation $\sigma_q (0 \leq \sigma_q \leq 2)$ to the projections of reference points $\boldsymbol{q}_j^i$.

Throughout this experiment, we evaluate the estimated distance parameter and its ground truth as error metrics. Here parameters with subscript g indicate ground truth data. The distance between $\boldsymbol{R}$ and $\boldsymbol{R}_g$, E_R, is defined as the Riemannian distance [29]:

$$E_R = \frac{1}{\sqrt{2}} \parallel \mathrm{Log}(\boldsymbol{R}^\top \boldsymbol{R}_g) \parallel_F \tag{10}$$

$$\mathrm{Log}\boldsymbol{R}' = \begin{cases} 0 & (\theta = 0), \\ \dfrac{\theta}{2\sin\theta}(\boldsymbol{R}' - \boldsymbol{R}'^\top) & (\theta \neq 0), \end{cases} \tag{11}$$

where $\theta = \cos^{-1}(\frac{tr\boldsymbol{R}'-1}{2})$. The difference between $\boldsymbol{t}$ and $\boldsymbol{t}_g$, E_t, is defined as an angle between two vectors:

$$E_t = \cos^{-1}(\boldsymbol{t}, \boldsymbol{t}_g). \tag{12}$$

In addition, we compare our proposal to Zhang's method [2] as a reference. In order to use this method, we set 30 reference points on the plane, which is a 5×6 grid pattern and the length of each reference point is 5 cm, in the area shared with C_0 and C_1. Notice these reference points are only for comparison and do not exist in the situation assumed.

2) Results With Synthesized Data: Fig. 6 shows the E_R and E_t of the proposed method and [2]. In each figure, the vertical axis shows the average value over 100 trials, and the horizontal axis denotes the standard deviation of noise.

From Fig. 6, we can observe that E_R and E_t decrease in the order of (a), (b), (c), and (d). While methods (a) and (b) use Algorithm1 in (Step2), methods (c) and (d) use Algorithm2. From this fact, we can say that Algorithm2 outperforms Algorithm1. This is due to the essential matrix step because it has been reported that the essential matrix computation is weak against observation noise [28]. While Algorithm2 only computes the essential matrix in (Step1), Algorithm1 computes it in (Step1) and use this result for computing the other essential matrix in (Step2). In addition, we can see that method (d) is equivalent to Zhang's method [2] from Fig. 6. This proves that our proposed method can estimate the extrinsic parameters without visible corresponding points with precision equivalent to that of [2] (assuming that visible corresponding points exist).

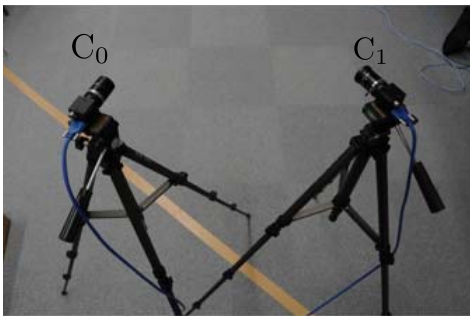

Fig. 7. Configuration for proposed method in indoor scene.

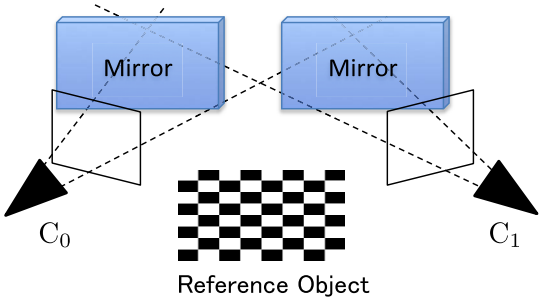

Fig. 8. Configuration for Mirror-Based approach [10] in indoor scene.

B. Real Data (Indoor Scene)

1) Experiment Environment: Fig. 7 provides an overview of the indoor scene used. We use Flea3 (PointGrey) for C_0 and C_1 (their resolution is 1600×1200), and use THETA S (RICOH) for X (its resolution is 5376×2688). The number of pairs of point correspondences between I_i and I_X range from 20 to 30, and the number of omnidirectional images N_X is 45. The intrinsic camera parameters $\boldsymbol{K}_0$ and $\boldsymbol{K}_1$ are estimated by Zhang's method [2] beforehand.

We compare our method to the mirror-based approach proposed by Takahashi *et al.* [10] as the reference. In order to use [10], we prepare a reference object, a 10×11 chessboard on which the length of each reference point is 2 cm, out of the cameras' shared FOV and set a mirror to allow observation of the reference object, as in Fig. 8.

TABLE II
E_R, E_t WITH REAL DATA (INDOOR SCENE)

Method	E_R	E_t
Takahashi et al.	0.0045	0.0368
(a)	0.0041	0.1717
(b)	0.0267	0.0249
(c)	0.0091	0.0388
(d)	0.0075	0.0256

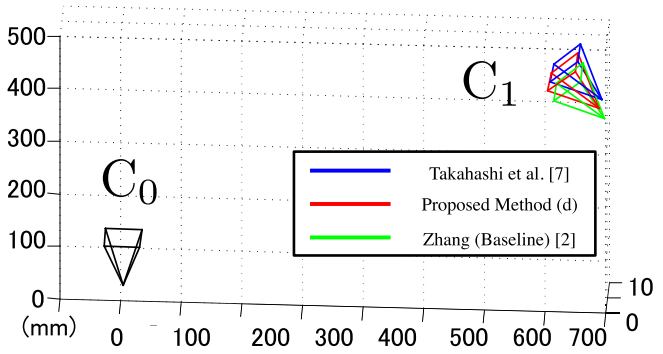

Fig. 9. Estimated positions of cameras C_1 estimated by the proposed method (d) (red), by [10] (blue) and by [2] (green). Notice that C_0 is located at $(0, 0, 0)^\top$.

In this evaluation, we use Zhang's method [2] as the baseline method. We set the chessboard in the shared FOV and estimate extrinsic parameters in [2]. To evaluate each method, we regard these parameters as the ground truth and use the same evaluation functions (10) and (12).

2) Results With Real Data: Table II quantitatively compares the parameters estimated by four variants of the proposed method with those of Takahashi's method [10]. We can see that the proposed method (especially method (d)) can estimate extrinsic parameters as precisely as [10]. Notice that the differences in the rotation matrix for the x-axis, y-axis, and z-axis are 2.77, 0.378, and 0.190 degrees, respectively ($E_R = 0.0075$). In addition, Fig. 9 shows the estimated positions of C_0 and C_1. Notice that the translation vector $\boldsymbol{t}$ estimated by the proposed method has an arbitrary scale, which we set $|\boldsymbol{t}| = |\boldsymbol{t}_g|$. From Fig. 9, we can see that the position of C_1 estimated by proposed method (d) almost matches those estimated by Takahashi's method [10] and Zhang's method [2]. From above, we can conclude that our method works properly in indoor environments.

Here, we collect pairwise corresponding points between each image I_i and each omnidirectional image I_X. The proposed method does not require any point corresponding multiple (more than two) cameras, which are needed to apply SfM with BA for estimating the extrinsic parameters of the cameras. It is not easy to find points corresponding with multiple cameras because color consistency between multiple cameras can not be easily preserved in general. It is one of the advantage of the proposed method that the extrinsic parameters between multiple cameras can be estimated using just pairwise corresponding points.

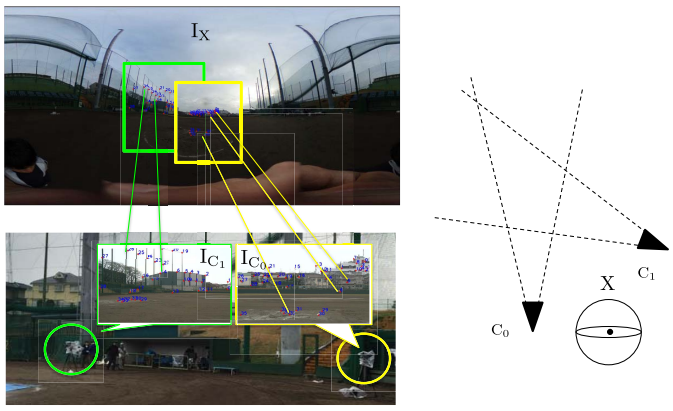

Fig. 10. Configuration for the proposed method in an outdoor scene. The left part shows the input images captured by C_0, C_1, and X. The Right part illustrates the geometric relation of the cameras. Notice that we do not have point correspondences between the images captured by C_0 and C_1.

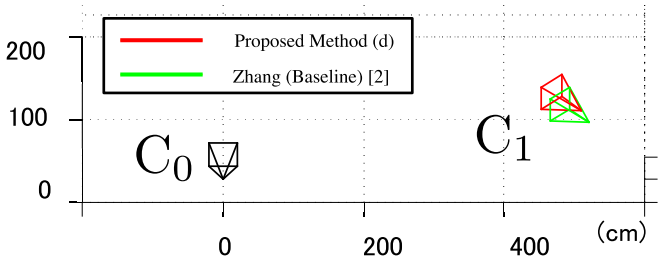

Fig. 11. Estimated positions of cameras C_1 estimated by the proposed method (d) (red) and by [2] (green). Notice that C_0 is located at $(0, 0, 0)^\top$.

C. Real Data (Outdoor Scene)

1) Experiment Environment: Fig. 10 shows an overview of the configuration of the outdoor scene, a baseball ground. We use an XDCam (Sony) for C_0 and C_1, with a resolution of 1280×720, and THETA S (RICOH) for X, with a resolution of 5376×2688. The number of pairs of point correspondences between I_i and I_X range from 30 to 35, and the number of omnidirectional images N_X is 26. The intrinsic camera parameters $\boldsymbol{K}_0$ and $\boldsymbol{K}_1$ are estimated by Zhang's method beforehand.

In this evaluation, we use Zhang's method [2] as the baseline method. We set the visual reference, an 11×14 checkerboard on which the length of each reference point is 14 cm, in the shared FOV and estimate the extrinsic parameters in [2]. To evaluate each method, we regard these parameters as the ground truth and use the same evaluation functions (10) and (12).

2) Result With Real Data: Table III quantitatively compares the parameters estimated by the four variants of the proposed method. We can see that proposed method (d)) can estimate extrinsic parameters with the same high precision as in the other experiments.

Fig. 11 shows the estimated positions of C_0 and C_1. Notice that we set $|\boldsymbol{t}| = |\boldsymbol{t}_g|$ for the same reason as in the indoor experiment. Fig. 11 shows that C_1 estimated by proposed method (d) almost matches that estimated by Zhang's method [2]. These results prove that the proposed

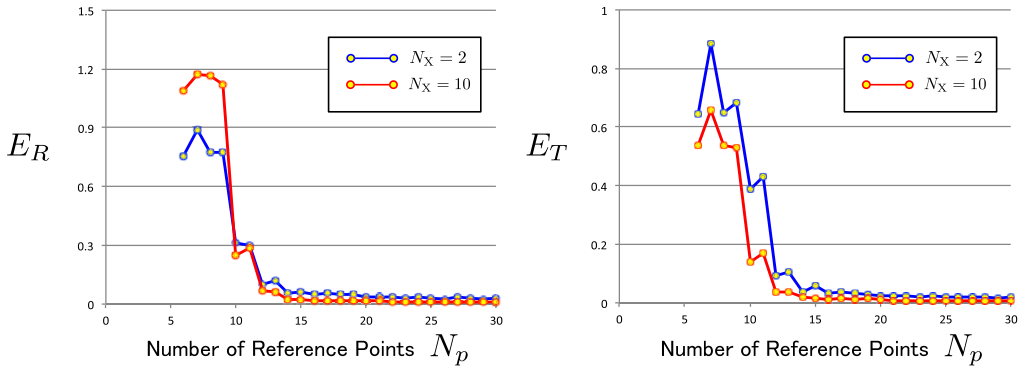

Fig. 12. Estimation error of each parameter in changing the number of reference points N_p.

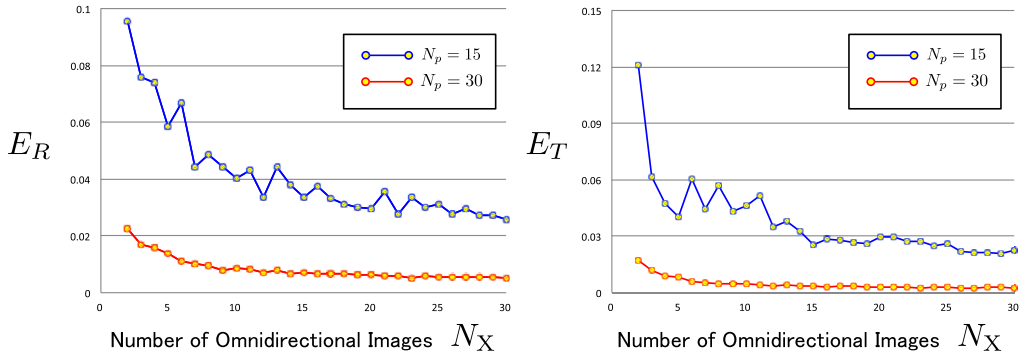

Fig. 13. Estimation error of each parameter in changing the number of omnidirectional images N_X.

TABLE III
E_R, E_t WITH REAL DATA (OUTDOOR SCENE)

Method	E_R	E_t
(a)	1.2990	0.8328
(b)	0.1053	0.1503
(c)	0.0267	0.0763
(d)	0.0072	0.0807

method works properly in expansive outdoor scenes, where the conventional method fails.

V. DISCUSSION

A. Impact of Number of Reference Points and Omnidirectional Cameras

In order to more closely examine our proposed method, we investigate the impact of the configuration, that is, the number of reference points and omnidirectional cameras. We focus here on variant method (d) in Table I. We use the same experiment environment as in Sec. IV-A. In the following evaluation, we add zero-mean Gaussian noise with a standard deviation $\sigma_q = 1$ to the input.

Fig. 12 shows the results gained while varying the number of reference points. We observe that the number of reference points has a strong impact on the estimation error of the proposed method, and we can say that more than 15 reference points should be used for stable extrinsic calibration.

Fig. 13 shows the results gained while varying the number of omnidirectional cameras. This figure proves that increasing the number of omnidirectional cameras also increases the accuracy of the proposed method. However, we can observe that the scale of error functions is smaller than in the case of changing the number of reference points. From these evaluations, we can say that the number of reference points should be increased for stable extrinsic calibration with high accuracy.

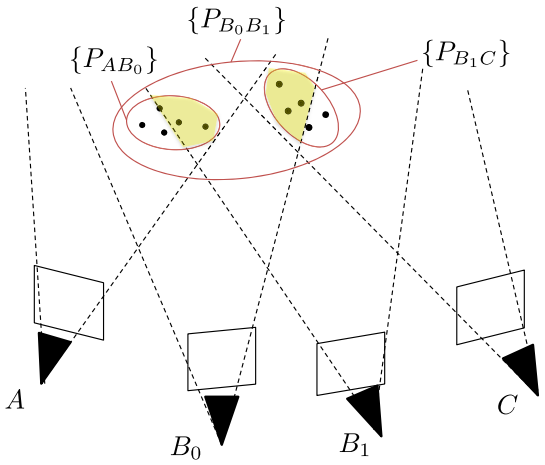

Fig. 14. Configuration for SfM approach. In order to calibrate A and C, points in the yellow area should be observed by $N'_X \geq 3$ cameras in order to transfer with a consistent scale.

B. Degenerate Case

Our algorithm does not work if it cannot compute an essential matrix. This happens if all reference points are on the same plane as reported in [1]. As for Algorithm1, we compute

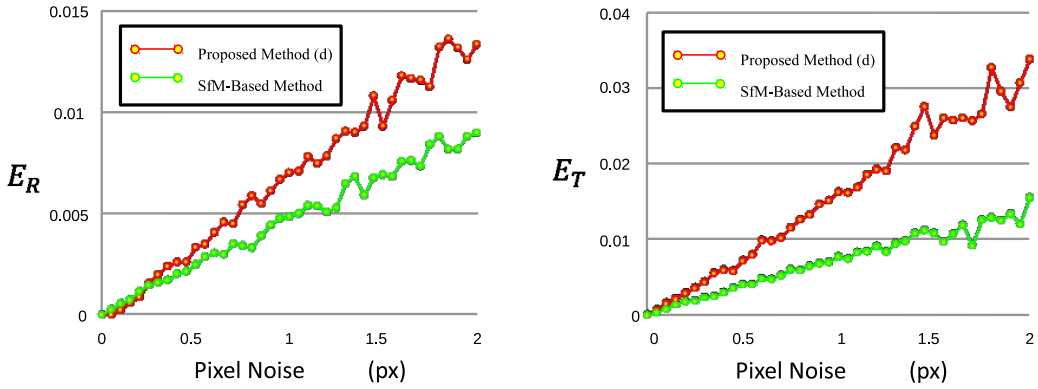

Fig. 15. Estimation error of proposed method and SfM.

the essential matrix in (Step1) and (Step2), so we should avoid the case where the positions of the omnidirectional cameras are on the same plane.

As for Algorithm2, when the vector connecting the center of two omnidirectional cameras parallels the vector connecting the center of C_0 and C_1, we cannot compute the translation vector $\boldsymbol{t}$ in (Step2), which yields the degenerate case, as reported in [28].

C. Comparison With Structure-From-Motion Based Methods

In order to calibrate multiple cameras with non-overlapping FOVs, some SfM based method can be adapted, such as [1]. Since these methods also utilize epipolar geometry, the proposed method has some similar properties with SfM-based method in terms of estimation precision. Fig. 15 shows the E_R and E_T of the proposed and SfM-based methods, and we can observe that the performance of the SfM-based method was not much different from that of our proposed method.

It is true the SfM-based method outperforms our proposed method. However, our method still has a significant advantage in that it can calibrate multiple cameras using only pairwise corresponding points between each camera and each omnidirectional camera, as we use only essential matrices between cameras. This means that we do not need to have any point that is a shared point for transferring with a consistent scale within multiple cameras, which is needed for applying a standard SfM with BA approach in the assumed configuration where the cameras to be calibrated do not have corresponding points.

In general, in order to perform a valid SfM/BA approach with a consistent scale in such a configuration, a point should be observed by camera groups comprising $N'_X \geq 3$ cameras for transferring with a consistent scale. For example, we assume that there are two cameras that have no corresponding points, C_0 and C_1, and two intermediate cameras, X_0 and X_1, as shown in Fig. 14. In this figure, we denote a 3D point sets observed by camera pairs C_0 and X_0 as $\{P_{C_0X_0}\}$. $\{P_{X_0X_1}\}$ and $\{P_{X_1C_1}\}$ also represent 3D point sets observed by camera pairs X_0 and X_1, and camera pairs X_1 and C_1. In this configuration, a part of $\{P_{C_0X_0}\}$ should be observed by (C_0, X_0, X_1) and also a part of $\{P_{X_1C_1}\}$ should be observed by (X_0, X_1, C_1) in order to calibrate with a consistent scale. It is difficult to satisfy this condition especially if there are some occlusions and/or significant differences of observation among the images captured by intermediate cameras.

On the other hand, our proposed method does not need to have any such corresponding points detected and matched within multiple cameras; rather, there only need to be corresponding points between two cameras (a camera and an omnidirectional camera), which are relatively easy to detect and match.

This advantage of our proposed method is that it is suitable for challenging scenes, such as when there are some occlusions or significant differences of observation among the images captured by multiple cameras. The correspondences are often inconsistent in such challenging scenes; therefore the SfM-based method will sometimes fail in estimating extrinsic parameters. The novelty of the proposed method and the advantage it provides compared with the SfM/BA approach lies in the derivation of extrinsic parameters, *i.e.*, not estimating extrinsic parameters through reconstructed 3D points but estimating them by utilizing the positions of omnidirectional cameras as corresponding points.

VI. Conclusion

In this paper, we proposed a novel algorithm that can calibrate multiple cameras scattered across a broad area, such that they do not have corresponding points in their shared FOVs. The key idea of our method is "using the position of an omnidirectional camera as the reference point." Based on this idea, we implement two types of algorithms for extrinsic calibration. In evaluations using synthesized data and real data, our method was found to be accurate.

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[2] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[3] F. Pagel, "Calibration of non-overlapping cameras in vehicles," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2010, pp. 1178–1183.
[4] P. Lébraly, C. Deymier, O. Ait-Aider, E. Royer, and M. Dhome, "Flexible extrinsic calibration of non-overlapping cameras using a planar mirror: Application to vision-based robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2010, pp. 5640–5647.
[5] J. A. Hesch, A. I. Mourikis, and S. I. Roumeliotis, "Mirror-based extrinsic camera calibration," in *Algorithmic Foundation of Robotics VIII*. Berlin, Germany: Springer, 2009, pp. 285–299.

[6] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, Nov. 2001.
[7] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 569–577, 2003.
[8] C. Lipski, F. Klose, and M. Magnor, "Correspondence and depth-image based rendering a hybrid approach for free-viewpoint video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 6, pp. 942–951, Jun. 2014.
[9] J. Y. Lee, H.-C. Wey, and D.-S. Park, "A fast and efficient multi-view depth image coding method based on temporal and inter-view correlations of texture images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 12, pp. 1859–1868, Dec. 2011.
[10] K. Takahashi, S. Nobuhara, and T. Matsuyama, "A new mirror-based extrinsic camera calibration using an orthogonality constraint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 1051–1058.
[11] R. Rodrigues, J. P. Barreto, and U. Nunes, "Camera pose estimation using images of planar mirror reflections," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 6314. Sep. 2010, pp. 382–395.
[12] Y. Caspi and M. Irani, "Aligning non-overlapping sequences," *Int. J. Comput. Vis.*, vol. 48, no. 1, pp. 39–51, Jun. 2002.
[13] S. Esquivel, F. Woelk, and R. Koch, "Calibration of a multi-camera rig from non-overlapping views," in *Proc. 29th DAGM Joint Pattern Recognit. Symp.*, vol. 4713. Sep. 2007, pp. 82–91.
[14] I. Kitahara, H. Saito, S. Akimichi, T. Ono, Y. Ohta, and T. Kanade, "Large-scale virtualized reality," in *Proc. Comput. Vis. Pattern Recognit., Tech. Sketches (CVPR)*, Dec. 2001, p. 4.
[15] R. K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys, "Simple calibration of non-overlapping cameras with a mirror," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–7.
[16] A. Agrawal, "Extrinsic camera calibration without a direct view using spherical mirror," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2368–2375.
[17] C. Nitschke, A. Nakazawa, and H. Takemura, "Display-camera calibration using eye reflections and geometry constraints," *Comput. Vis. Image Understand.*, vol. 115, no. 6, pp. 835–853, Jun. 2011.
[18] J.-S. Kim, M. Hwangbo, and T. Kanade, "Motion estimation using multiple non-overlapping cameras for small unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2008, pp. 3076–3081.
[19] P. Lébraly, E. Royer, O. Ait-Aider, C. Deymier, and M. Dhome, "Fast calibration of embedded non-overlapping cameras," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2011, pp. 221–227.
[20] G. Carrera, A. Angeli, and A. J. Davison, "SLAM-based automatic extrinsic calibration of a multi-camera rig," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2011, pp. 2652–2659.
[21] E. Ataer-Cansizoglu, Y. Taguchi, S. Ramalingam, and Y. Miki, "Calibration of non-overlapping cameras using an external SLAM system," in *Proc. 2nd Int. Conf. 3D Vis. (3DV)*, vol. 1. Dec. 2014, pp. 509–516.
[22] T. Strauß, J. Ziegler, and J. Beck, "Calibrating multiple cameras with non-overlapping views using coded checkerboard targets," in *Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2623–2628.
[23] A. Rahimi, B. Dunagan, and T. Darrell, "Simultaneous calibration and tracking with a network of non-overlapping sensors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2004, pp. I-187–I-194.
[24] Z. Liu, X. Wei, and G. Zhang, "External parameter calibration of widely distributed vision sensors with non-overlapping fields of view," *Opt. Lasers Eng.*, vol. 51, no. 6, pp. 643–650, 2013.
[25] W. Zou and S. Li, "Calibration of nonoverlapping in-vehicle cameras with laser pointers," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1348–1359, Jun. 2015.
[26] Z. Liu, G. Zhang, Z. Wei, and J. Sun, "Novel calibration method for non-overlapping multiple vision sensors based on 1D target," *Opt. Lasers Eng.*, vol. 49, no. 4, pp. 570–577, Apr. 2011.
[27] D. Farin, J. Han, and P. H. N. de With, "Fast camera calibration for the analysis of sport sequences," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2005, pp. 482–485.
[28] J. C. Bazin, C. Demonceaux, P. Vasseur, and I. S. Kweon, "Motion estimation by decoupling rotation and translation in catadioptric vision," *Comput. Vis. Image Understand.*, vol. 114, no. 2, pp. 254–273, Feb. 2010.
[29] M. Moakher, "Means and averaging in the group of rotations," *SIAM J. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 1–16, 2002.

Shogo Miyata received the B.Sc. and M.Sc. degrees in information and computer science from Keio University, Japan, in 2015 and 2017, respectively. He is currently employed by SRD, a game software company in Japan.

Hideo Saito (S'90–M'92–SM'09) received the Ph.D. degree in electrical engineering from Keio University, Japan, in 1992. Since then, he has been on the Faculty of Science and Technology, Keio University. From 1997 to 1999, he was with the Virtualized Reality Project, Robotics Institute, Carnegie Mellon University, as a Visiting Researcher. Since 2006, he has been a Full Professor with the Department of Information and Computer Science, Keio University. His research interests include computer vision and pattern recognition, and their applications to augmented reality, virtual reality, and human robotics interaction. He is a fellow of the Institute of Electronics, Information and Communication Engineers and Virual Reality Society of Japan. He has received awards, including the Best Paper Award in 3DSA2010, the Best Paper Award in VSMM2010, the Honorable Mention for Best Short Paper in the IEEE VR in 2011, the Most Influential Paper over the Decade Award of MVA in 2000, the Jon Campbell Best Paper Prize in IMVIP2014, and a Best Paper in Electronic Imaging in 2016. His recent activities for academic conferences, include being the Program Chair of ACCV2014 and ISMAR2016, and the General Chair of ISMAR2015.

Kosuke Takahashi received the B.Sc. degree in engineering and the M.Sc. degree in informatics from Kyoto University, Japan, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Researcher with NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation. His research interest includes computer vision. He received the Best Open Source Code Award Second Prize in CVPR 2012.

Dan Mikami received the B.E. and M.E. degrees from Keio University, Kanagawa, Japan, in 2000 and 2002, respectively, and the Ph.D. degree from Tsukuba University in 2012. He has been with Nippon Telegraph and Telephone Corporation since 2002. His current research activities include computer vision and information technologies for enhancing sport performance. He received the Meeting on Image Recognition and Understanding Excellent Paper Award in 2009, the IEICE Best Papar Award in 2010, the IEICE KIYASU-Zen'iti, Award in 2010, and the IPSJ SIG-CDS Excellent Paper Award in 2013.

Mariko Isogawa received the B.S. and M.S. degrees from Osaka University, Japan, in 2011 and 2013, respectively. She has been with Nippon Telegraph and Telephone Corporation since 2013. Her research interests include multimedia content handling.

Akira Kojima received the B.E. and M.E. degrees in mathematical engineering and information physics from The University of Tokyo in 1988 and 1990, respectively. He is currently the Manager with the Media Innovation Business Department, NTT TechnoCross Corporation. Since joining NTT Group, Nippon Telegraph and Telephone Corporation, in 1990, he has been involved in research and development on video database, digital library, multimedia information retrieval, video surveillance, and high-reality visual communication. He is a member of the Institute of Electronics, Information and Communication Engineers, the Institute of Image Electronics Engineers of Japan, and the ACM.