Accurate and Practical Calibration of Multiple Pan-Tilt-Zoom Cameras for Live Broadcasts

MASANORI KANO1, HIDEHIKO OKUBO2, MASAKI TAKAHASHI1, KENSUKE IKEYA1, KENSUKE HISATOMI1 (Member, IEEE), AND TOMOYUKI MISHINA1

1Science and Technology Research Laboratories, Japan Broadcasting Corporation (NHK), Tokyo 157-8510, Japan
2Broadcast Engineering Department, Japan Broadcasting Corporation (NHK), Tokyo 150-8001, Japan

Corresponding author: Masanori Kano (kanou.m-gc@nhk.or.jp)

ABSTRACT We aim to realize novel services for live broadcasts, including visualizing the 3D position of a subject, using multiple pan-tilt-zoom (PTZ) cameras. For this purpose, a camera calibration technique with high accuracy over a wide pan-tilt-zoom range and high operability is required. Herein, we propose an accurate and practical method to calibrate multiple PTZ cameras for live broadcasts. This method models the PTZ camera system and accurately pre-estimates unknown parameters in the model. After the pre-estimation, accurate camera parameters can be calculated in real time using the pan-tilt angles and zoom values acquired from the sensors in the camera system. The proposed method accurately estimates the unknown parameters for a high-magnification zoom lens by making repeated initial guesses and nonlinear optimizations for high accuracy, and consists of a two-stage camera calibration for operability. We demonstrated the effectiveness of the proposed method by conducting experiments on computer graphics data and real data. We also applied it to a live sports graphics system.

INDEX TERMS Camera calibration, live broadcast, multiple cameras, pan-tilt-zoom camera.

I. INTRODUCTION

We aim to realize novel services for live broadcasts by using computer graphics (CG) and computer vision (CV) technologies. For example, we can capture a subject using multiple cameras, analyze its 3D position, and synthesize CG in the images of live broadcasts, such as those of sports events. Such effects will make live TV programs easier to follow.

To calculate the 3D position of a subject, it is necessary to capture the subject with multiple cameras. Additionally, there are two ways of capturing a moving subject in a wide space, such as a sports stadium, at high resolution. One way is to use many fixed cameras, each of which captures the subject with a static angle of view and covers part of the space. The other uses several pan-tilt-zoom (PTZ) cameras that can track the subject. Implementing the first method, which requires time and effort to set up many fixed cameras, is difficult in live sports broadcasts, as the preparation time is not sufficient. Moreover, if the number of cameras is reduced and the subject is captured with a wider angle of view, the resolution of the image is significantly lowered. However, a small number of PTZ cameras can efficiently capture a wide space at high resolution. Therefore, we have incorporated multiple PTZ cameras in this study.

The cameras must be calibrated to analyze the 3D positions of subjects in images and synthesize CGs. Camera calibration is a technique for estimating camera parameters, which consist of extrinsic parameters representing the position and orientation of the camera and intrinsic parameters representing the characteristics of the camera and the lens. A fixed camera has to be pre-calibrated only once because its parameters do not change until it is moved. Conversely, because a PTZ camera moves, its parameters change constantly. Hence, a calibration method for a fixed camera cannot be applied to a PTZ camera. Additionally, to take advantage of a PTZ camera’s characteristics, it must be calibrated over a wide pan-tilt-zoom range.

We thus established four requirements for our camera calibration:

- Robustness: cameras can always be calibrated without fail.
- Real time: camera parameters can be obtained during operation in real time.
- Accuracy: 3D position analysis and CG synthesis can be conducted over a wide pan-tilt-zoom range without causing discomfort to the viewer.
II. RELATED WORK

The camera calibration techniques of Tsai [1] and Zhang [2] use a known calibration object. In Tsai’s method, a camera is calibrated using both known 2D points in an image and 3D points in the space, whereas in Zhang’s method, a camera is calibrated by capturing a planar calibration pattern several times through a process that changes the position and orientation of the camera or the calibration pattern. Ueshiba and Tomita [3] extended Zhang’s method to work with a multi-camera system and improved the calibration accuracy and stability. Willson proposed a calibration method for a camera with a zoom lens [4]. This method is based on Tsai’s camera model and models a camera with a zoom lens as a polynomial function. Although these methods are effective for fixed or zoom cameras, it is difficult to apply them to PTZ cameras.

Recently, techniques utilizing natural feature points, such as Simultaneous Localization and Mapping (SLAM) [5] and Structure from Motion (SfM) [6], have been actively studied. These methods have high operability because no calibration object is required. SLAM, which has been actively studied in the field of robotics, estimates the position and orientation of the camera in real time. However, most SLAM methods assume that the intrinsic parameters are known and cannot handle zoom changes. On the other hand, SfM is a technique for estimating 3D structures of the scene and camera parameters from 2D image sequences simultaneously. Although SfM can also estimate intrinsic parameters, it cannot do so in real time because it has many variables to estimate. Both techniques lack robustness due to their dependence on captured images. Additionally, because matching between images is necessary, the cameras are required to capture similar images. Therefore, the camera arrangement is restricted; for instance, the cameras cannot be sparsely placed throughout a wide area.

PTZ cameras are often used in surveillance systems because they can direct attention to particular events in a scene [7]. Several calibration approaches have been proposed for these systems [8]–[10]. However, in many cases, PTZ cameras can be modeled using a simple motion model that assumes the pan-tilt rotation axes pass through the optical center of the camera [11]. We use a more detailed camera model to improve the accuracy of camera calibration. Davis et al. proposed an accurate camera model [12], which only supports pan and tilt movements, and not zoom.

Camera calibration methods exist for live broadcasts [13], [14]. Okubo et al. proposed a method [14] for CG synthesis in a virtual studio using a single PTZ camera. Their method models a PTZ camera and estimates the unknown parameters of the model beforehand by using calibration data. After the estimation, the camera parameters can be unfailingly obtained during operation in real time from the sensor values. The method enables accurate CG synthesis for the calibration pattern used for estimation with a 13× zoom lens, but it also has some issues. Although it uses an initial guess and nonlinear optimization similar to the general camera calibration method, the initial guess method, which is an important but difficult step in optimization, is not specified. Moreover, accuracy is only verified over a narrow pan-tilt range; it must be experimentally evaluated over a wider range as well. These issues are overcome in the proposed method.

III. CAMERA MODEL

We model the PTZ camera using the method reported in [14]. The model considers a whole camera system, where a camera equipped with a zoom lens is mounted on the pan-tilt camera platform shown in Fig. 1. The camera is regarded as a perspective projection.

Let us now consider the intrinsic parameters; note that focus is not taken into consideration. In a zoom lens, the intrinsic parameters change according to the zoom settings. Subsequently, the counter value of the zoom lens measured by the sensor is defined as the zoom value \( n \). Intrinsic parameters \( \kappa \), which depend on \( n \), are expressed as a vector defined by the following equation:

\[
\kappa(n) = [f(n), a(n), c_u(n), c_v(n), q_1(n), q_2(n), q_3(n), q_4(n)]^T
\]

where \( f \), \( a \), \([c_u, c_v]^T\), and \([q_1, q_2, q_3, q_4]^T\) are focal length, aspect ratio, principal point, and lens distortion coefficients,
respectively. Equation (1) without the lens distortion coefficients is expressed in matrix notation as a matrix of intrinsic parameters:

\[
K(n) = \begin{bmatrix} f(n) & 0 & c_u(n) \\ 0 & a(n)f(n) & c_v(n) \\ 0 & 0 & 1 \end{bmatrix}.
\] (2)

The extrinsic parameters are modeled in consideration of the characteristics of a PTZ camera. Generally, extrinsic parameters represent the relationship of position and orientation between two coordinate systems. The extrinsic parameters \( \mu \) are expressed as follows:

\[
\mu = [t_x, t_y, t_z, \theta_x, \theta_y, \theta_z]^T,
\] (3)

where the translation and angle vectors between the two coordinate systems are \( t = [t_x, t_y, t_z]^T \) and \( \theta = [\theta_x, \theta_y, \theta_z]^T \), respectively. When transforming from one coordinate system, \( \Sigma_A \), to another, \( \Sigma_B \), these extrinsic parameters are expressed as \( B \mu_A \). Their rigid-body transformation, \( B M_A \), is described as

\[
B M_A = \begin{bmatrix} B R_A \\ 0^T \\ 1 \end{bmatrix},
\] (4)

where \( B t \) is the translation vector in the coordinate system \( \Sigma_B \), and \( B R_A \) is the rotation matrix obtained from the angle vector. The superscript of a translation or position vector indicates its coordinate system, and the subscript and superscript of the matrix indicate the transformation between the coordinate systems. Using the above notation, the extrinsic parameters of a PTZ camera can be modeled (see Fig. 1).

In this camera model, the general extrinsic parameters are decomposed and modeled by considering the characteristics of the pan-tilt camera platform. Supposing that the pan-tilt axes are orthogonal at one point in the platform, this point is defined as a rotational center. Let the coordinate system at the rotational center represent the platform coordinate system \( \Sigma_{CP} \) when both pan and tilt angles are zero. The extrinsic parameters from the world coordinate system \( \Sigma_W \) to \( \Sigma_{CP} \) are denoted as \( C F \mu_W \). Let the coordinate system after pan-tilt motions at the rotational center represent the pan-tilt coordinate system \( \Sigma_{PT} \). The extrinsic parameters from \( \Sigma_{CP} \) to \( \Sigma_{PT} \) are \( C F \mu_{CP} \). Subsequently, because the pan-tilt axes cross at one point, the translation vector of \( C F \mu_{CP} \) becomes \( 0 \), and the orientation consists of only pan and tilt angles. Therefore, \( C F \mu_{CP} \) (or \( C F M_{CP} \)) is calculated from only the pan and tilt angles obtained by the angle sensors. Let the coordinate system of the camera with the camera’s optical center as its origin be the camera coordinate system \( \Sigma_C \). The extrinsic parameters from \( \Sigma_{PT} \) to \( \Sigma_C \) are \( C \mu_{PT} \). Because the position of the camera coordinate system changes with the zoom value, \( C \mu_{PT} \) is a function of \( n \). The element of this parameter, which is used as a function of the zoom value, depends on the design of the camera model. Here, we assume that only the translation vector in the direction of the optical axis, namely the z-axis direction, depends on the zoom value.

\[
C \mu_{PT}(n) = [t_x^C, t_y^C, t_z^C(n), \theta_x^C, \theta_y^C, \theta_z^C]^T = [\xi n^C, t_z^C(n)]^T.
\] (5)

The extrinsic parameters \( C \mu_w \) from \( \Sigma_W \) to \( \Sigma_C \), which are equivalent to \( C M_w \), are obtained by

\[
C M_w(p, t, n) = C M_{PT}(n) C M_{CP}(p, t) C M_w
\] (6)

where \( p, t, \) and \( n \) are pan angle, tilt angle, and zoom value, respectively. Equation (6) indicates that if we estimate two extrinsic parameters \( C \mu_w \) and \( C \mu_{PT} \), then we can calculate extrinsic parameter \( C \mu_w \) from only the pan-tilt angles and zoom value, namely the sensor values, acquired from the sensors.

The unknown parameters to be estimated in this camera model are the intrinsic parameter \( \kappa(n) \), and extrinsic parameters \( C F \mu_w \) and \( C \mu_{PT} \). We refer to \( \kappa(n)^T = (t_x^C(n), t_y^C(n))^T \) zoom-independent parameters and \( [\xi n^C, t_z^C(n)]^T \) zoom-dependent parameters. Because estimating parameters with all the zoom values is difficult, we estimate them with the sampled zoom values \( n = N_k \) \((k = 1, 2, \ldots, K) \). Subsequently, the zoom-dependent parameters between the sampled ones are interpolated. We define these as camera model parameters (CMPs) \( \phi \), as follows:

\[
\phi = [\kappa(N_1)^T, \ldots, \kappa(N_K)^T, t_z^C(N_1), \ldots, t_z^C(N_K), \xi n^C, C \mu_{WT}]^T
\] (7)

Let the CMPs of the \( i \)-th camera, \( C_i \), be \( \phi_i \). After the parameters are estimated, the camera parameters can be robustly obtained from only the sensor values in real time. In addition, the more accurately the parameters can be estimated, the more accurately the camera parameters are obtained. Once the CMPs are estimated, the intrinsic and extrinsic parameters, \( \kappa(n) \) and \( C \mu_{PT}(n) \), respectively, do not change unless the camera, lens, or camera platform is replaced. This property is used in the on-location camera calibration described later.

Let us consider the projection of a point in 3D space to a camera image. The point \( Q \), the coordinate of which in \( \Sigma_W \) is \( X = [X, Y, Z]^T \), is projected onto the image as follows:

\[
su = P(p, t, n) \tilde{X},
\]

\[
P(p, t, n) = K(n) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} C M_w(p, t, n).
\] (8)
where $s, u = [u, v]^T$, and $P$ are the scale factor, image coordinates, and projection matrix, respectively. In addition, $\hat{\cdot}$ indicates the homogeneous coordinates of its vector.

Because the above-described projection model involves an ideal lens without distortion, we address the projection of a point in consideration of lens distortion. The coordinate of a 3D point $\mathbf{w}X$ in $\Sigma_W$ is transformed into $\mathbf{C}W\mathbf{w}X$ in $\Sigma_C$ by $\mathbf{C}W$. The normalized image coordinate $x = [x, y]^T = [\mathbf{C}X]_ZW = \mathbf{C}[\mathbf{X}, \mathbf{Y}, \mathbf{C}]^T$, where $\mathbf{C}X = [\mathbf{X}, \mathbf{Y}, \mathbf{Z}]^T$. The normalized image coordinate $x_d = [x_d, y_d]^T$ including lens distortion is as follows:

$$
\begin{bmatrix}
  x_d \\
  y_d
\end{bmatrix} =
\begin{bmatrix}
  (1 + q_1r^2 + q_2r^4)x + 2q_3xy + q_4(r^2 + 2x^2) \\
  (1 + q_1r^2 + q_2r^4)y + q_3(r^2 + 2y^2) + 2q_4xy
\end{bmatrix},
$$

(9)

where $r^2 = x^2 + y^2$. The image coordinates $u_d = [u_d, u_d]^T$ including lens distortion are obtained by multiplying $x_d$ by $K$:

$$
\tilde{u}_d =
\begin{bmatrix}
  u_d \\
  v_d \\
  1
\end{bmatrix} = K(n)
\begin{bmatrix}
  x_d \\
  y_d \\
  1
\end{bmatrix}
$$

(10)

IV. CALIBRATION ALGORITHM

The camera calibration method is divided into two phases: an initial camera calibration (ICC) performed beforehand and an on-location camera calibration (OCC) performed at the filming location. The CMPs are estimated and interpolated using ICC, and the parameters are saved in a database. Subsequently, only the parameters dependent on the filming location must be updated by employing OCC. This strategy enables accurate camera calibration to be conducted quickly and reduces the amount of adjustment required at the filming location. To estimate the parameters, calibration data are obtained by capturing calibration patterns. Once the CMPs are estimated, the calibration patterns become redundant during operation.

Let us consider the type and arrangement of calibration patterns and the coordinate system. Each calibration pattern is a planar object on which known feature points are printed, e.g., a checkerboard pattern. For a high-magnification zoom lens, it is difficult to capture calibration patterns properly with only one type of pattern over the entire zoom range. Therefore, we use calibration patterns of various sizes. A large calibration pattern is captured over the entire wide zoom range, and a small one is captured in the tele zoom range. Additionally, to improve the accuracy of the camera calibration in a wide space, multiple calibration patterns are placed throughout that space. The calibration patterns are freely arranged; their positions and orientations may be unknown because they are estimated simultaneously with the CMPs. The calibration pattern has a coordinate system, which is known as the calibration pattern coordinate system $\Sigma_P$. Then, the relationship of the pose and orientation between the world and calibration pattern coordinate system is expressed in the same way as the extrinsic parameters. The extrinsic parameters from $\Sigma_W$ to $\Sigma_P$ are represented as $^P\mu_W$. For simplicity, we express $v = ^P\mu_W$ (see Fig. 1). In the case of the $j$-th calibration pattern $P_j$ ($j = 1, 2, \ldots, J$), the coordinate system is $\Sigma_{P_j}$, and the extrinsic parameters from $\Sigma_W$ to $\Sigma_{P_j}$ are $v_j$ or $^P\mu_{W_j}$.

The calibration data consist of images captured with the calibration patterns in them and sensor values for estimating the CMPs. Each camera captures several shootable calibration patterns at each sampled zoom value for multiple poses, i.e., at different pan and tilt angles. We refer to the number of changes as the number of poses. The captured images and sensor values are then paired and stored. As described above, whether the calibration pattern can be appropriately captured depends on the size of the calibration pattern and the zoom value. Consequently, a table is created that consists of the type of calibration pattern and sampled zoom value. Fig. 2 shows an example of the calibration data, and Fig. 2 (a) shows an example of the calibration data structure. In this example, there are four calibration patterns, i.e., $P_1$ to $P_4$, and four sampled zoom values, i.e., $N_1$ to $N_4$. Here, we assume that the size of the calibration pattern decreases as the number increases from $P_1$ to $P_4$, and the sampled zoom value changes from the wide side to the tele side as the number increases from $N_1$ to $N_4$. The cells with data are colored gray, and the cells without data are colored white. Each gray cell includes the number of poses, and there are multiple data captured with various poses, as shown in Figs. 2 (b) and (c). As a result, all images and sensor values captured with various combinations of cameras, calibration patterns, sampled zoom values, and poses become calibration data.

Next, we explain ICC and OCC. To accurately estimate CMPs, it is advantageous to have more calibration data. Hence, we want more calibration patterns, sampled zoom values, and poses. However, a large amount of calibration data is difficult to capture at the filming location because time is limited, and many calibration patterns cannot be arranged freely. Each calibration involves capturing calibration data and estimating parameters, but the most time and effort is invested in capturing the calibration data. Therefore, it is important to reduce the number of captured calibration data at the filming location. To achieve high accuracy with little time and effort, the method is divided into two phases: ICC and OCC.

When camera $C_i$ ($i = 1, 2, \ldots, I, I \geq 2$) captures a feature point $m$ ($m = 1, 2, \ldots, M_j$) on the calibration pattern $P_j$ ($j = 1, 2, \ldots, J$) with zoom $k$ ($k = 1, 2, \ldots, K_j$) and with pose $l$ ($l = 1, 2, \ldots, L_{ij}$) and the 3D coordinates of the point are projected onto the image, we denote the observed image coordinates as $\mathbf{u}_{ijklm}$ and estimated image coordinates as $\mathbf{u}_{ijkm}$. $\omega_{jk}$ is used as an indicator variable, where $\omega_{jk} = 1$ if $v_j$ and $[\mathbf{e}(N_j)_k]_0^T, t_{ij}^c(N_j)_k$ have been estimated, and $\omega_{jk} = 0$ otherwise. The subscripts in the parameter and coordinate systems represent camera number $i$, calibration pattern number $j$, zoom number $k$, pose number $l$, and feature point number $m$. 153996
A. INITIAL CAMERA CALIBRATION (ICC)
The purpose of ICC is to estimate all CMPs. Although a large amount of calibration data must be captured over time for accuracy, the amount of work required is not an issue because it is conducted in advance of arriving at the filming location. We arrange many calibration patterns in a wide space and capture calibration data with many sampled zoom values. The parameters to be estimated in ICC are as follows:

\[ \eta = [\phi_1^T, \ldots, \phi_I^T, \nu_1^T, \ldots, \nu_J^T]^T. \]  

The CMPs of all cameras and the extrinsic parameters of all calibration patterns are estimated. Note that we can freely set the world coordinate system \( \Sigma_W \), which is the basis of a space. By matching this with an arbitrary calibration pattern coordinate system, we can eliminate six parameters—the translation and orientation—from (11). We call this the basic calibration pattern \( P_1 \) and consider this case from here on.

Next, we discuss the CMP estimation. The proposed method repeatedly performs initial guesses and nonlinear optimizations for estimating the parameters while adding calibration data at each iteration. Fig. 3 (a) shows a flowchart, while Figs. 3 (b) and (c) show examples of the change in calibration data used and of the parameters estimated in each step with written numbers. The basic estimation estimates some of the parameters; then, the calibration pattern expansion estimates extrinsic parameters of the calibration pattern apart from the basic one; and the zoom expansion estimates zoom-dependent parameters apart from that of the basic zoom value. The basic zoom value, i.e., the sampled zoom value used for the basic estimation, may be any sampled zoom value. Hereafter, we assume that it is \( N_1 (k = 1) \). The details of each step in ICC are as follows.

1) FEATURE POINT EXTRACTION
The feature points of the calibration patterns are extracted from all of the images of the calibration data.

2) APPLICATION OF CALIBRATION METHOD FOR FIXED CAMERAS
The calibration method for a fixed camera is applied only to the calibration data acquired by capturing the basic calibration pattern with the basic zoom value. Zhang’s method [2]
can be used here because each camera captures the calibration pattern in multiple poses. The camera parameters—intrinsic parameters $\kappa_i(N_1)$ and extrinsic parameters $C_i\mu_W$—of each camera are estimated (see Fig. 4).

3) INITIAL GUESS FOR BASIC ESTIMATION
This process estimates the CMPs only at the basic zoom value by using the result of the previous one. In particular, $t^C(N_1)$, $\xi^C$, and $C^P\mu_W$ are estimated. The extrinsic parameters $C_i\mu_W$ can be decomposed into $C_p\mu_{PT}(N_1)$ and $CP\mu_W$ by applying the method described in [15] to the obtained extrinsic parameters $C_i\mu_W$ in multiple poses (see Fig. 4).

4) OPTIMIZATION FOR BASIC ESTIMATION
After the initial values of the partial CMPs have been estimated, we optimize them for each camera. We conduct nonlinear optimization to refine the parameters $\tilde{\phi}_i$ by minimizing the projection error using the Levenberg–Marquardt algorithm [16], [17]. This process is performed for each camera. The equation for camera $C_i$ is as follows:

$$\tilde{\phi}_i = \arg \min_{\phi_i} \sum_{l=1}^{L_{ik}} \sum_{m=1}^{M_j} \| u_{ijklm} - \hat{u}_{ijklm}(\phi_i) \|^2, \tag{12}$$

where $j = 1$, $k = 1$, and $\hat{\cdot}$ indicates estimated parameters.

5) INITIAL GUESS FOR CALIBRATION PATTERN EXPANSION
The extrinsic parameters of the calibration pattern apart from the basic one are estimated using the estimated CMPs at that instant. Parameters $\nu_j$ are obtained by calculating the 3D positions of the feature points of $P_j$ and by using the correspondence between the 3D positions of the feature points of the basic calibration pattern $P_1$ and target calibration patterns $P_j$. First, we calculate the 3D coordinate $X_{jm}$ of the feature point $m$ of $P_j$ in $\Sigma_W$ (see Fig. 4). A projection matrix $P_{ijkl}$ is obtained using $\phi_i$ for all cameras, zooms, and poses using (6) and (8). Because the undistorted image coordinates $u_{ijklm}$ of the feature points on $P_j$ are known, multiple equations of (8) hold as follows:

$$\begin{bmatrix}
P_{ijkl} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & 0 \\
\vdots & \cdots & \ddots & \vdots \\
0 & \cdots & 0 & \tilde{u}_{ijklm}
\end{bmatrix}$$

These equations are set up for all 3D positions of the feature points of the target calibration pattern $X_{jm}$ is obtained by solving them [18]. Next, because the correspondence of the feature points between the basic and target calibration pattern is known, the extrinsic parameters $\nu_j$ from $\Sigma_{P_1}$ to $\Sigma_{P_j}$ are calculated using the position and orientation estimation method reported in [19]–[21]. Note that the differences in size and number of feature points between calibration patterns must be adjusted in advance. Although the maximum number of possible simultaneous equations is described in (13), the equations that have un-estimated parameters are actually eliminated because they cannot hold.

6) INITIAL GUESS FOR ZOOM EXPANSION
The zoom-dependent parameters of each zoom value other than the basic zoom value are estimated using the parameters estimated at that time. These parameters are calculated using CMPs that have already been estimated, the 3D position $X$ of the calibration pattern with the estimated extrinsic parameters, and their distorted image coordinates $u$. Equation (8) represents these relationships. In this equation, the unknown zoom-dependent parameters other than lens distortion coefficients are $[f, a, c_u, c_v, \xi^C]^T$, and we estimate them. Therefore, we define $\alpha$, $\beta$, and $\gamma$ as follows:

$$\begin{bmatrix}
\alpha \\
\beta \\
\gamma + t^C_1
\end{bmatrix} = \begin{bmatrix}
a_{11} & a_{12} & a_{13} & b_1 + t^C_1 \\
a_{21} & a_{22} & a_{23} & b_2 + t^C_2 \\
a_{31} & a_{32} & a_{33} & b_3 + t^C_3 \\
0 & 0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
X^T \\
Y \\
Z \\
1
\end{bmatrix} \begin{bmatrix}
C_{p_{ll}}(n)^{PT}M_{p_{ij}(i,p)}^{CP}M_{W} \\
\tilde{\chi}
\end{bmatrix}. \tag{14}$$

By considering the estimated calibration patterns, poses, and feature points for which calibration data exist, the following equation can be derived from (8) and (14):

$$\begin{bmatrix}
\alpha_1 & 0 & \gamma_1 & 0 \\
0 & \beta_1 & 0 & \gamma_1 \\
0 & 0 & \gamma_1 & 0 \\
\alpha_E & 0 & \gamma_E & 0
\end{bmatrix} \begin{bmatrix}
f \\
af \\
c_u \\
\gamma v_{1,1}
\end{bmatrix} = \begin{bmatrix}
\gamma v_{1,1}h_{1,1} \\
\gamma v_{1,1} \gamma_{V,1,1} \\
\gamma v_{1,1} c_{u} \\
\gamma v_{1,1} \gamma_{v,E,F}
\end{bmatrix} \begin{bmatrix}
\gamma v_{1,1} \gamma_{V,1,1} \\
\gamma v_{1,1} c_{u} \\
\gamma v_{1,1} \gamma_{v,E,F}
\end{bmatrix}. \tag{15}$$

where $E$ is a number depending on the numbers of calibration patterns and poses, and $F$ is a number depending on the
The optimized parameters $\hat{\eta}$ are estimated. The final $\eta$ is estimated after the last optimization of Fig. 3 (a). CMPs $\phi_i$ in the ICC are determined as a result.

8) PARAMETER INTERPOLATION

Interpolation is required for the zoom-dependent parameters. We use spline interpolation.

B. ON-LOCATION CAMERA CALIBRATION (OCC)

The purpose of OCC is to estimate part of the CMPs with fewer calibration data to reduce the time and labor required at the filming location. Therefore, the parameters that change with camera relocation are estimated. Because OCC is similar to ICC, we only describe the different points. The intrinsic parameters $\kappa_i$ and extrinsic parameters $C_i\mu_{PT_i}$ have already been determined by ICC, and they do not change. We estimate only the extrinsic parameters $C_{\phi_i}\mu_{W'}$ when transitioning from the new world coordinate system $\Sigma_{W'}$ to the camera platform coordinate system $\Sigma_{CP_i}$ of each camera. The new basic calibration pattern is $P_1'$, and the new basic zoom value is $N_1'$. In the OCC, we estimate $\phi_i' = [C_{\phi_i}\mu_{W'}]_1^T$. Additionally, in this case, because multiple calibration patterns $P'_j (j = 1, 2, \ldots, J')$ are used, the parameters to be estimated are as follows:

$$\eta' = \left[\phi_1'^T, \ldots, \phi_{J'}^T, v_1', \ldots, v_{J'}'\right]^T.$$  \hspace{1cm} (17)

Fig. 5 (a) shows a flowchart of the process. Additionally, Figs. 5 (b) and (c) show examples of the change in calibration data used and of the parameters estimated in each step with written numbers, respectively. By reducing the numbers of calibration patterns, sampled zoom values, or poses, the calibration data can be reduced in comparison to the amount of ICC data. The sampled zoom value is defined as $n = N'_k (k = 1, 2, \ldots, K'_j)$. $'$ indicates that the values have been changed from ICC to OCC values. The details of these steps are presented below.

1) APPLICATION OF CALIBRATION METHOD FOR FIXED CAMERAS

A calibration method for a fixed camera is only applied to the calibration data when capturing a new basic calibration pattern with a new basic zoom value.

2) INITIAL GUESS FOR BASIC ESTIMATION

Because $C_i\mu_{W'}, PT_i\mu_{CP_i}$, and $C_i\mu_{PT_i}$ are known, $C_{\phi_i}\mu_{W'}$ is calculated using (6).

3) OPTIMIZATION FOR BASIC ESTIMATION

The parameters $\phi_i'$ are optimized nonlinearly by minimizing the projection error of each camera. For camera $C_i$, the equation is as follows:

$$\phi_i' = \arg\min_{\phi_i'} \sum_{k=1}^{K'_j} \sum_{l=1}^{L'_j} \|u_{ijklm} - \hat{u}_{ijklm}(\phi_i')\|^2,$$  \hspace{1cm} (18)

where $j = 1$ and $k = 1$. 

VOLUME 8, 2020

M. Kano et al.: Accurate and Practical Calibration of Multiple PTZ Cameras for Live Broadcasts

153999
4) CALIBRATION DATA EXPANSION
Because the zoom-dependent parameters are already known, calibration data are added.

5) OPTIMIZATION
The only difference from the ICC is that the parameter follows (17).

\[
\hat{\eta}' = \arg\min_{\eta'} \sum_{i=1}^{J'} \sum_{j=1}^{J'} \sum_{k=1}^{M'_j} \sum_{m=1}^{M'_j} \omega_{ijkm} \left\| \hat{u}_{ijkm}(\eta') - \hat{u}_{ijkm}(\eta) \right\|^2.
\]

(19)

As a result, the necessary parameters \( \phi'_i \) are estimated by OCC.

V. EXPERIMENTS
The main purpose of the experiments was to assess the accuracy and operability of the proposed camera calibration method, as the method guarantees real-timeness and robustness. The effectiveness of the method was confirmed through experiments using CG data and real data.

A. CG DATA
Two cameras and four calibration patterns were placed in the CG space, as illustrated in Fig. 6. The resolution, principal point, and lens distortion coefficients of the camera were HD (1920 × 1080 pixels), half of the resolution, and [0, 0, 0, 0], respectively. The zoom lens had 10× magnification, and the zoom had a range of 0–999.

1) ESTIMATION ACCURACY OF CMPs
To investigate estimation accuracy, we estimated CMPs using ICC and compared them with the ground truth (GT) values. We used two PTZ cameras, colored in red in Fig. 6, in ICC. The calibration data were acquired according to Table 1, where WE and TE indicate the wide and tele ends of the zoom lens, respectively. The number of sampled zoom values was seven, and the number of poses was a maximum of five. The maximum number of poses was five because calibration data that could not be extracted from the image were not used. Fig. 7 and Table 2 present the results. All CMPs were estimated accurately. Although the relative errors of \( t_x \) and \( t_y \) in (5)—the same as \( t'_x \) and \( t'_y \) in (5)—appear large; this is because these scales are smaller than the others. We also considered when these scales were larger (5: \( t'_x = [-209.9482, 189.2410] \), \( t'_y = [-212.9196, 277.4071] \). The others were the same.). In the case of larger scales, the relative error was improved. This result shows that the proposed method can accurately estimate CMPs.

2) PROJECTION ACCURACY AND EFFECTIVENESS OF CAMERA MODEL
We investigated the projection accuracy of ICC and the effect of the camera model used (detailed camera model described in Section III).

a: PROJECTION ACCURACY
We investigated projection accuracy by projecting 3D points onto an image. The points were placed in a spherical shape
TABLE 2. Estimation results of zoom-independent parameters.

| C1 | \(e_{\mu_x}(x^C)\) | \(e_{\mu_y}(y^C)\) | \(e_{\mu_z}(z^C)\) | \(\theta_x [\text{deg}]\) | \(\theta_y [\text{deg}]\) | \(\theta_z [\text{deg}]\) |
|----|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| ICC | -1.015 | 0.389 | 0.032 | 1.000 | 2.000 | 3.000 |
| OCC | -2.746 | 0.906 | 0.321 | 1.089 | 1.987 | 2.999 |

TABLE 3. Calibration data structure in OCC for CG data experiments.

| \(N_x\) | \(P_x(50)\) | \(P_x(60)\) | \(P_x(40)\) |
|-------|---------|---------|---------|
| 0     | 5       | 5       | 5       |
| 500   | 5       | 5       | 5       |
| 999   | 5       | 5       | 5       |

with a radius of 5000 mm; therefore it surrounded the two cameras (see Fig. 6). We set the pan and tilt angle in the range of \(-81–80\) degrees in 5 degree increments and the zoom value in the range of 0–999 in increments of 50. In each trial, we measured the projection error in the image coordinate system. The results are shown in Fig. 8 (a). In all trials, the mean, minimum, and maximum errors were 1.24, 0.24, and 5.81 pixels, respectively. In Fig. 8 (a), the points representing the error are plotted as the mean error by collecting the pan-tilt angles every 10 degrees and zoom value every 200. The error tended to increase as the pan-tilt angles moved to the edge where the calibration patterns were not placed. We confirmed that the projection error is small over a wide pan-tilt-zoom range, which means that the camera calibration accuracy is high.

**b: EFFECTIVENESS OF DETAILED CAMERA MODEL**

We estimated the CMPs with the simple camera model using ICC and measured the projection accuracy. In the simple camera model, \(t_x^C\), \(t_y^C\), and \(t_z^C\) in CMPs (see (5)) are set to zero in the optimization phase. The results are shown in Fig. 8(b). In all trials, the mean, minimum, and maximum errors were 24.77, 1.46, and 140.83 pixels, respectively. Clearly, the error of the simple camera model was larger than that of the detailed camera model. Furthermore, in the case of larger scales with similar conditions as V-A1, the difference was more significant. The mean error was 1.50 pixels for the detailed camera model and 94.79 pixels for the simple one. The error for the detailed camera model was almost unchanged but that for the simple camera model was larger than the smaller scales case. This result shows that the detailed camera model is effective for improving accuracy. Moreover, the longer the distance between the rotational and optical centers, the greater the effect is. Therefore, the proposed method can be applied to camera systems with longer distances.

**3) EFFECTIVENESS OF OCC**

We confirmed the ability of OCC to maintain the calibration accuracy with a small amount of calibration data. OCC was performed after relocating cameras (cameras colored in cyan in Fig. 6) and acquiring calibration data according to Table 3.

The results of the mean projection error are shown in Fig. 9. Fig. 9 (b) shows the mean error from collecting the zoom value for each increment of 50. In all trials, the mean, minimum, and maximum errors were 1.95, 0.32, and 9.91 pixels, respectively. Although the mean error of OCC increased by 0.71 pixels compared with that of ICC, it maintained high accuracy over a wide pan-tilt-zoom range. This indicates that we can perform OCC if we acquire a small amount of calibration data at the filming location. Therefore, OCC helps improve operability.

We investigated the case of performing ICC using the calibration data for OCC. The results of the projection error are shown in Fig. 10. Fig. 10 (a) shows that the error increases...
TABLE 4. PTZ camera specifications.

| Video format | Resolution | Frame rate | Focal length | Magnification | Pan and tilt | Zoom value |
|--------------|------------|------------|--------------|---------------|-------------|------------|
| HD (1920 × 1080) | 59.94 Hz | 4.7–94 mm | 20 | 800, 000 / 360 [pulse/deg] | 0(W5)–1025(TB) |

TABLE 5. Calibration data structure in ICC for real data experiments. The sizes of the calibration patterns are in units of mm.

| Nm | PA(200) | PA(250) | PA(300) | PA(350) | PA(400) | PA(450) |
|----|---------|---------|---------|---------|---------|---------|
| Nm = 60 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 80 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 100 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 200 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 300 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 400 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 500 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 600 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 700 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 800 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 900 | 5 | 5 | 5 | 5 | 5 | 5 |
| Nm = 1000 | 5 | 5 | 5 | 5 | 5 | 5 |

FIGURE 12. Experimental environment. The upper and lower panoramic pictures show the environment of the parameter estimation and accuracy verification, respectively. (a) ICC. (b) OCC.

FIGURE 13. Parts of the estimation results of the zoom-dependent parameters. (a) Focal length f. (b) Extrinsic parameter tCz.

In the experiments, we used four PTZ cameras, as shown in Fig. 11. The specifications of the PTZ camera are listed in Table 4. The camera and its lens were a MEC-4000 CHZ-
1360 made by Carina System Co., Ltd. The pan-tilt platform was custom made.

1) INITIAL CAMERA CALIBRATION (ICC)

a: ESTIMATION OF CMPs

We estimated the CMPs of all cameras using ICC. The experimental environment is shown in Fig. 12 (a). The number of calibration patterns was eight, the number of sampled zoom values was 22, and the number of poses was a maximum of five. Table 5 shows the calibration data structure. When the CMPs were estimated, the mean reprojection error was 1.18 pixels. This implies that the CMPs were estimated accurately because the error converged to a small value. Fig. 13 shows the parts of the zoom-dependent parameters (interpolated) in the CMPs. Fig. 13 (a) shows that the focal length $f$ naturally changes with the zoom value. However, in the zoomed-in area of Fig. 13 (b), $t_{c}^z$ changed somewhat unnaturally. We confirmed that the feature points in the images of the calibration data used for estimation were accurately projected.

b: ACCURACY OF CAMERA PARAMETERS

We verified the accuracy of the camera parameters when operating the PTZ cameras freely, i.e., with an arbitrary setting of the pan-tilt-zoom range. We marked several points on an image—indicated by $\times$—and drew epipolar lines on the other images. The results are shown in Fig. 15, where the enlarged image is $51 \times 51$ pixels. By operating the cameras over a wide pan and zoom range, it was confirmed that the epipolar lines were accurately drawn. However, the epipolar lines deviated by several pixels from the corresponding points in some data. This occurred when the panning or tilting was large, and the zoom was on the tele side at the same time. This was because the camera model was not perfect, and the estimation error of the parameters was conspicuous, as shown in Fig. 13 (b).

2) ON-LOCATION CAMERA CALIBRATION (OCC)

a: ESTIMATION OF CMPs

We estimated the parameters $CP_{W}$ of all the cameras by using OCC. The experimental environment is shown in Fig. 12 (b). Here, we examine two cases: OCC (1)
and OCC (2). The purpose of OCC is to estimate parameters as accurately and simply as possible. In these two cases, the number of calibration patterns and the sampled zoom values were varied, and differences in accuracy and operability were examined. Tables 6 and 7 show the calibration data structures. The calibration data were common.

When the parameters were estimated, the mean reprojection error was 3.43 pixels in OCC (1) and 2.74 pixels in OCC (2). Although the error increased more than in ICC camera calibration because of cumulative error, these results imply that the parameters were still estimated accurately. This is because the error is small. It is apparent that OCC (1), which has more calibration data, has more errors because its parameters must adjust for a large amount of calibration data. However, this indicates the possibility of achieving high accuracy on average over a wider pan-tilt-zoom range. We confirmed that the projected feature points on the images of the calibration data used for estimation were accurate.

b: ACCURACY OF CAMERA PARAMETERS

We verified the accuracy of the camera parameters when operating the PTZ cameras freely. We used epipolar lines for the accuracy verification similar to the experiment described in Section V-B1.b. As shown in Fig. 15, the accuracy was generally high. However, it tended to decrease when the pan-tilt angles were large, and the zoom was on the tele side (see Fig. 15(c)). The accuracy of OCC (1) was higher than

TABLE 7. Calibration data structure of OCC (2).

| N | P1(100) | P2(40) | P3(100) | P4(40) |
|---|---------|--------|---------|--------|
| 6 | 5       | 5      | 5       | 5      |
that of OCC (2) on the whole. This is because OCC (1) had more calibration data than OCC (2). In OCC (2), the accuracy declined in Fig. 15(c) because calibration data with a zoom range greater than 750 ($n > 750$) were not used. However, the accuracy was generally equal to that of OCC (1) when the zoom range was 750 or below. Thus, capturing calibration data in the required range only is efficient. The final accuracy defined by the requirements (see Section I) is described in the next section.

**c: OPERABILITY**

Let us consider the time taken by ICC and OCC. When acquiring the calibration data, we captured the data 490 times with each camera in ICC. However, data were captured 125 times in OCC (1) and 50 times in OCC (2). Therefore, by using OCC (especially OCC (2)), we could decrease the number of captures to approximately 1/10. Additionally, less calibration data help reduce calculation time for parameter estimation.

**VI. APPLICATION**

We incorporated the proposed camera calibration method into a live sports graphics system that we developed. The system calculates the 3D position of a ball from only images [22] and synthesizes a CG ball in images captured by an arbitrary camera in real time. An overview of the process is illustrated in Fig. 16 (a). The proposed method provides the system with camera parameters to calculate the 3D position of the ball and synthesizes CG. We used four PTZ cameras, arranged as shown in Fig. 16 (b). The camera system and its arrangement were the same as in the experiment described in Section V-B2. The cameras were calibrated by OCC. Although the conditions were approximately the same as those for OCC (2) in Section V-B2, the amount of data was slightly less. A green CG was synthesized on a volleyball. It was found that CG balls could be accurately displayed in the images despite the pan-tilt-zoom range being wide (Fig. 16 (c)). This proves that the proposed method could calculate camera parameters accurately in real time. Moreover, the total delay from capture to output was approximately 0.3 seconds, and all cameras were operated by a single cameraman [23].

We exhibited the system at the NHK STRL Open House 2017 [24]. Images from the event are shown in Fig. 17. The mark (NHK logo) and trajectory were accurately synthesized. We set up the system easily using OCC. The demonstration proved that the camera calibration method has sufficient accuracy and operability for a live sports graphics system.

**VII. CONCLUSION**

We proposed an accurate and practical camera calibration method for multiple PTZ cameras for live broadcasts of events, such as a sports event. The proposed method enables the acquisition of accurate camera parameters over a wide pan-tilt-zoom range in real time. In addition, for high operability at the filming location, we separated the initial and on-location camera calibration. We experimentally verified the effectiveness of our method and demonstrated a live sports graphics system that incorporates it. In the future, we plan to apply our method to live broadcasts.

**REFERENCES**

[1] R. Tsai, “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses,” IEEE J. Robot. Autom., vol. RA-3, no. 4, pp. 323–344, Aug. 1987.
[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] T. Ueshiba and F. Tomita, “Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices,” in Proc. 9th IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2003, pp. 966–973.
[4] R. G. Willson, “Modeling and calibration of automated zoom lenses,” Proc. SPIE, vol. 2350, pp. 170–186, Oct. 1994.
[5] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” in Proc. 6th IEEE ACM Int. Symp. Mixed Augmented Reality, Nov. 2007, pp. 225–234.
[6] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: Exploring photo collections in 3D,” ACM Trans. Graph., vol. 25, no. 3, pp. 835–846, 2006.
[7] H. Aghajan and A. Cavallaro, Multi-Camera Networks: Principles and Applications. New York, NY, USA: Academic, 2009, pp. 189–190.
[8] A. Abrams and R. Pless, “Web-accessible geographic integration and calibration of webcams,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 9, no. 1, pp. 1–20, Feb. 2013.
[9] S. N. Sinha and M. Pollefeys, “Towards calibrating a pan-tilt-zoom camera network,” in Proc. ECCV Workshop Omnidirectional Vis. Camera Netw., 2004, pp. 42–54.
[10] Z. Wu and R. J. Radke, “Keeping a pan-tilt-zoom camera calibrated,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1994–2007, Aug. 2013.
[11] K. Ikeuchi, Computer Vision: A Reference Guide. New York, NY, USA: Springer, 2014, pp. 581–586.
[12] J. Davis and X. Chen, “Calibrating pan-tilt cameras in wide-area surveillance networks,” in Proc. Int. Conf. Comput. Vis. (ICCV), 2003, pp. 144–149.
[13] H. Okubo, Y. Yamanouchi, H. Mitsumine, T. Fukaya, and S. Inoue, “Accurate camera calibration method specialized for virtual studios,” Proc. SPIE, vol. 6804, pp. 131–142, Feb. 2008.
[14] H. Okubo, H. Mitsumine, T. Fukaya, and Y. Yamanouchi, “A study on high-power zoom lens calibration method for virtual studios,” (in Japanese), Inst. Electron., Inf. Commun. Eng. (IEICE), Tokyo, Japan, IEICE Tech. Rep. MVE, 2008, pp. 61–66, vol. 108, no. 226.
[15] H. Zhuang, Z. S. Roth, and R. Sadjikha, “Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form AX=XB,” IEEE Trans. Robot. Autom., vol. 10, no. 4, pp. 549–554, Aug. 1994.
[16] K. Levenberg, “A method for the solution of certain non-linear problems in least squares,” Quart. Appl. Math., vol. 2, no. 2, pp. 164–168, 1944.
[17] D. W. Marquardt, “An algorithm for least-squares estimation of nonlinear parameters,” J. Soc. Ind. Appl. Math., vol. 11, no. 2, pp. 431–441, Jun. 1963.
[18] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003, pp. 588–596.
[19] K. Arun, T. Huang, and S. D. Blostein, “Least-squares fitting of two 3-D point sets,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 5, pp. 543–576, Sep. 1987.
[20] K. Kanatani, “Analysis of 3-D rotation fitting,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 16, no. 5, pp. 543–549, May 1994.
[21] S. Umetani, “Least-squares estimation of transformation parameters between two point patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, no. 4, pp. 376–380, Apr. 1991.
[22] M. Takahashi, K. Ikeya, M. Kano, H. Ookubo, and T. Mishina, “Robust volleyball tracking system using multi-view cameras,” in Proc. Int. Conf. Pattern Recognit. (ICPR), 2016, vol. 4, no. 4, pp. 2740–2745.
[23] K. Ikeya and Y. Iwadate, “Bullet time using multi-viewpoint robotic camera system,” ITE Trans. Media Technol. Appl., vol. 4, no. 4, pp. 349–362, 2016.
[24] NHK STRL Open House 2017. Accessed: Jun. 11, 2020. [Online]. Available: https://www.nhk.or.jp/stli/open2017/tenji/4_e.html
M. Kano et al.: Accurate and Practical Calibration of Multiple PTZ Cameras for Live Broadcasts

MASANORI KANO received the B.S. and M.S. degrees from Sophia University, Tokyo, Japan, in 2007 and 2009, respectively. In 2009, he joined Japan Broadcasting Corporation (NHK), Tokyo, where he has been working with the Science and Technology Research Laboratories since 2013. His research interests include camera calibration, 3D image processing, and 3D display.

KENSUKE IKEYA received the B.S. degree in engineering from the Tokyo University of Agriculture and Technology, Tokyo, Japan, in 2004, the M.S. degree in engineering from The University of Electro-Communications, Tokyo, in 2006, and the Ph.D. degree in engineering from the Tokyo Institute of Technology, Tokyo, in 2018. He joined Japan Broadcasting Corporation (NHK), Tokyo, in 2006. Since 2009, he has been involved in research on multi-viewpoint robotic cameras and their applications with NHK Science and Technology Research Laboratories. He is currently working with the Planning and Coordination Division.

MASAKI TAKAHASHI received the B.S. and M.S. degrees in engineering from Keio University, Japan, in 1997 and 1999, respectively, and the Ph.D. degree from the Graduate University of Advanced Studies, Japan, in 2012. He joined Japan Broadcasting Corporation (NHK) in 1999, where he has been working with the Science and Technology Research Laboratories, since 2002. He was a Visiting Researcher with the University of British Columbia in 2012. His research interests include visual object tracking and content production technology.

KENSUKE HISATOMI (Member, IEEE) received the B.E. and M.E. degrees in applied physics from Waseda University, Shinjuku, Japan, in 1997 and 1999, respectively, and the Ph.D. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2016. He joined Japan Broadcasting Corporation (NHK), Tokyo, in 1999. From 2003 to 2007, he was with the Engineering Administration Department, NHK, where he was involved in the development of digital broadcasting. He has been with the Science and Technology Research Laboratories, NHK, since 2007. He was a Visiting Researcher with Surrey University, Guildford, U.K., from 2010 to 2011. His current research interests include augmented reality and virtual reality.

HIDEHIKO OKUBO received the B.S. degree from Waseda University, Japan, in 1993. He joined Japan Broadcasting Corporation (NHK) in 1993. After working with the Science and Technology Research Laboratories as a Researcher of image processing and computer graphics, since 2018, he has been working with the Technical Operation and Engineering Division as a VFX Supervisor and a Production Pipeline Engineer.

TOMOYUKI MISHINA received the B.S. and M.S. degrees from the Tokyo University of Science, Japan, in 1987 and 1989, respectively, and the Ph.D. degree from the Tokyo Institute of Technology, Japan, in 2007. In 1989, he joined Japan Broadcasting Corporation (NHK), Tokyo, Japan. Since 1992, he has been working on threedimensional imaging systems with the Science and Technology Research Laboratories, NHK.

KENSUKE HISATOMI (Member, IEEE) received the B.E. and M.E. degrees in applied physics from Waseda University, Shinjuku, Japan, in 1997 and 1999, respectively, and the Ph.D. degree in information science and technology from the University of Tokyo, Tokyo, Japan, in 2016. He joined Japan Broadcasting Corporation (NHK), Tokyo, in 1999. From 2003 to 2007, he was with the Engineering Administration Department, NHK, where he was involved in the development of digital broadcasting. He has been with the Science and Technology Research Laboratories, NHK, since 2007. He was a Visiting Researcher with Surrey University, Guildford, U.K., from 2010 to 2011. His current research interests include augmented reality and virtual reality.