Estimating Camera Pose using Trajectories Generated by Pan-tilt Motion

Hiroto Nagayoshi
Central Research Laboratory, Hitachi Ltd., Japan
hiroto.nagayoshi.wy@hitachi.com

Marc Pollefeys
ETH Zürich, Switzerland
marc.pollefeys@inf.ethz.ch

Abstract

A novel method for auto-calibration of a PTZ (pan-tilt-zoom) camera network is proposed. The key idea on which it is based is to use pan-tilt motions generated by PTZ cameras themselves as calibration patterns. Generating and observing the pan-tilt motions of each camera makes it possible to estimate whole relative camera poses in a camera network. Cameras first observe circular trajectories of a marker set on another camera performing tilt motions with various pan angles. Although finding correspondences between marker points captured by the cameras is cumbersome due to network delays, that of between circular trajectories is easy. It is thus possible to use several hundred points in a circular trajectory to estimate its normal vectors and centers with high accuracy. Imposing geometric constraints on them makes it possible to eliminate ambiguities of them and obtain a unique relative camera pose directly. A novel refinement process to minimize differences between modeled pan-tilt motions and three or more circular trajectories is then performed. Experiments using synthetic data confirmed that the refinement process improved camera pose estimation accuracy in comparison to the direct estimation. An experiment using real data showed that the proposed method works properly for real cameras.

1. Introduction

Distributed PTZ (pan-tilt-zoom) camera networks are increasingly used nowadays. The density of camera coverage has also been increasing. That is, multiple cameras can often observe the same spatial location. However, to take full advantage of this capability, it is critical that the relative poses of all the cameras in the network is known. Obtaining the relative poses can be a cumbersome and expensive process when traditional surveying or pattern-based calibration techniques are used.

Promising approaches to obtain the relative poses are based only on feature points detected from images of natural scenes. In the case of these approaches, neither calibration patterns nor agents to perform certain actions are needed.

The approaches can save human loads when the number of cameras is large. However, indoor scenes, such as in large offices and corridors, often have very limited textures. Moreover, recent surveillance cameras are set in factories and plants. Feature points in those scenes are not detected from textures; instead, they are detected from patterns generated by combinations of pipes, cables, and other 3D structures. Since those patterns are not projective invariants, the approaches based on feature points do not work properly in the case of such indoor scenes.

Therefore, a novel method for estimating relative camera poses based on pan-tilt motions generated by a PTZ camera is proposed. Indeed, the only parts of the environment that can be directly controlled by the camera network are the cameras themselves, so they are used as calibration objects. This approach allows the network to estimate relative camera poses autonomously without any separate agents to perform certain actions in the environment.

To use the cameras as calibration objects, two prerequisites should be satisfied. The first is high image resolution and high zoom ratio, because the sizes of the cameras are small such as 20 (cm). This prerequisite is practical because modern PTZ cameras have Full HD resolution and more than 20 times optical zoom ratio. The second is that the positions of the cameras are adjusted so as to observe each other. For factories and plants in the planning stage, it is possible to apply a specification that satisfies the second prerequisite to newly-introduced camera networks.

The proposed method needs a marker set on a camera and tracks it to obtain circular trajectories generated by pan-tilt motions of the camera. Correspondences at the level of circular trajectories were required (i.e. not point-wise). Even though unknown network and image-encoding delays occur, keeping the correspondences is easier than that of the point-wise method. For instance, keeping the correspondences is possible by confirming that the marker is not moving for a few seconds and has been set to the initial pan-tilt state of each pan-tilt motion after the state is set to the cameras. The marker is tracked and several-hundred marker points are generated from one pan-tilt motion. As a result, the camera pose estimation can be more accurate. The cir-
circular trajectories in 3D space are observed as ellipses in images. A normal vector and the center of the circle can be estimated from the ellipses; however, the estimation has a two-fold ambiguity. This ambiguity can be avoided by incorporating a geometric constraint into the estimation; thereby, a unique relative camera pose is obtained from two rotations directly. A refinement process to minimize differences between modeled pan-tilt motions and observed three or more circular trajectories is then performed. The process is also a feature of the proposed approach because higher accuracy can be achieved by performing more pan-tilt motions.

2. Related work

As for methods for camera pose estimation using circles, those using concentric circles [2, 4] and coplanar circles [1, 5, 7, 8] have been proposed. However, PTZ cameras cannot generate those kind of circles. Wang et al. proposed a method using only one circle; however, it requires a known center of the circle [9], which is not known in the case of a PTZ camera. Wu et al. proposed a method that uses parallel circles [10]; in other words, two circles are on different parallel planes. This condition is satisfied when pan motions with different tilt angles are performed. However, cameras are usually set at the same height, so pan rotations are observed as lines by another camera.

The proposed camera pose estimation method is based on that proposed by Rahman et al. [6], which uses two arbitrary located circles. To resolve the ambiguity of the normal vector and the center of the circle, they assumed another condition that two circles are placed with certain distance. Pan motions satisfy this condition, however, they are not observed as circles as mentioned in the previous paragraph. Therefore, another geometrical condition is proposed. Although their method needs only two circles, more circles can be generated by more pan-tilt motions. The refinement process is applied to achieve high estimation accuracy.

3. Overall method

The basic idea on which the proposed method is based, is shown in Fig. 1. A reference camera performs tilt motions, and their trajectories are used to estimate relative positions and directions of target cameras. For the sake of simplicity, target cameras are referred to as cameras. The reason that tilt motions were chosen is two-fold. The first is that radii of tilt rotations can be kept constant. The absolute radii values can be got from the specification of the PTZ camera and used to calculate the absolute distance between the cameras.

The second is that pan motions are observed as lines in the usual case; that is, the cameras are set at similar height.

The flow of the proposed method is shown in Fig. 2. First, a marker that is set on the reference camera gives tilt motions with different pan angles. The trajectories of the marker are captured as ellipses in images of the target cameras. In this paper, it is assumed that the cameras are normalized and the ellipses are normalized. Hyper renormalization is used for ellipse fitting to get ellipse parameters because of its high accuracy, especially in the case that only a part of an ellipse is observed [3]. For circle-pose estimation that is to estimate the normal vector and the center of the circle, the method proposed by Chen et al. [1] is applied. However, two-fold ambiguity of the estimation still has to be resolved. Imposing geometric constraints on the estimation makes it possible to resolve the ambiguity and obtain a unique relative camera pose directly. A refinement process to minimize differences between modeled pan-tilt motions and three or more circular trajectories is then performed.

4. Circle-pose estimation and its geometrical interpretation

4.1. Circle-pose estimation

A circle in 3D space is projected onto an image plane as an ellipse. According to Chen et al. [1], the normal vector and the center of the circle can be estimated from the projected ellipse. To make this paper self-contained, their estimation technique is briefly reviewed as follows. The ellipse is denoted using a $3 \times 3$ matrix $Q_{vi}$ as $(x, y, 1)Q_{vi}(x, y, 1)^T = 0$, where subscripts $v$ and $i$ are indexes of the cameras and circles, respectively. Namely, $Q_{vi}$ is an ellipse that is observed by the $v$-th camera and corresponds to $C_{vi}$ that refers to a circular trajectory generated by $i$-th tilt motion in $v$-th camera’s coordinates. Assume that $C_{vi}$ is moved by a rotation $R_{c,vi}$ and becomes $C_{c,vi}$ that is parallel to the image plane. As a result, another ellipse, $Q_{c,vi}$ is obtained as a projection of the moved circle. Since $C_{c,vi}$ is parallel to the image plane, $Q_{c,vi}$ becomes a circle. The relation between $Q_{vi}$ and $Q_{c,vi}$ can then be
transfers \( R_y,vi \) transfers \( C_{x,vi} \) to a circle whose projection becomes ellipse \( \Lambda_{vi} \) (Fig. 3 (ii)). Since \( V_{vi} \) has ambiguity concerning its sign, Eq. (4) can be rewritten as
\[
R_{c,vi}(\alpha_{vi}) = V_{vi} S_{1,vi} R_y,vi R_z(-\alpha_{vi}),
\]
where \( S_{1,vi} \) expresses ambiguities concerning signs of rows in \( V_{vi} \). The ambiguities expressed by \( S_1 \) and \( S_2 \) are fully expressed by \( S_{1,vi} \). Note that without loss of generality, it is assumed that the sign of each eigenvector of \( V_{vi} \) is adjusted so that \( V_{vi} \) represents a rotation. Since it is enough to consider the case of \( z > 0 \), it is assumed that the (3,3)-th of \( V_{vi} \) is positive to avoid a rotation over the x-y plane. Similarly, a diagonal matrix \( S_{1,vi} \) is a rotation matrix with a positive (3,3)-th element as follows:
\[
S_{1,vi} = \begin{pmatrix} s_{1,vi} & 0 & 0 \\ 0 & s_{1,vi} & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]
5. Relative camera pose estimation

5.1. Direct method

It is known that relative camera poses between the target cameras can be estimated if two circles are observed by them[6]. However, according to Eqs. (13) and (14), a two-fold ambiguity, besides the sign of the normal vector, remains. When two circles are captured by two cameras, eight-fold ambiguity must be removed. Rhamann et al.[6] proposed a method for resolving this ambiguity as a discrete optimization problem. However, when the circles are the rotated ones sharing a common axis of rotation, the ambiguity cannot be resolved even when more than two circles are given. In this paper, the axis corresponds to an axis of pan motion. Generally, possible relative positions of a camera that is viewing a circle form a cone whose axis is orthogonal to the circle and passes through the center of the circle. When plural viewed circles are given, the possible camera positions are the intersections of the cones. Then the circles have a common axis of rotation, these intersections lie symmetrically on both sides of the plane that the axes of the cones span.

Therefore, another constraint must be used to remove the ambiguity. Originally, the ambiguity of the circle-pose estimation comes from that there are two cases to obtain a circle from an intersection between a cone and an plane. Therefore, it is easily removed when the closer side of the circle to the camera is determined. That is determined by putting a marker on a flat surface of the PTZ camera. Since normal vectors of the marker and a plane spanned by the circular trajectories are orthogonal in that case, the side of the circular trajectory from which the marker is detected is closer to the camera. The condition that removes the ambiguity is thus given as

\[(p_{vij} - c_{v}(s_{1,v}, s_{2,v}))^T n_{v}(s_{1,v}, s_{2,v}) < 0, \quad j = 1...n_j, \quad (15)\]

where \(p_{vij}\) is the \(j\)-th observed point of the \(i\)-th circle, observed by the \(v\)-th camera, and \(n_j\) denotes number of observed points. Since all the ambiguities are eliminated, it is possible to obtain two corresponding vectors and two corresponding positions in both camera. Thereby, the relative pose between target cameras can be estimated.

5.2. Optimization method

A refinement process to achieving higher estimation accuracy by using more than two rotations is proposed as follows. The method is based on the idea that the original circle can be expressed in two different ways. The first way is based on the circle-pose estimation. The second way is based on a pan-tilt rotation model. The refinement process aims to minimize the difference between the two expressions. This minimization is possible because the geometrical interpretation of the circle-pose estimation was clarified as mentioned in Sec. 4.2.

Pan-tilt model: Modern PTZ cameras such as Canon VB-H41 can be assumed to have orthogonal axes of pan motion and tilt motion. A simple pan-tilt model, as shown in Fig. 4, was then used to parametrize tilt rotations with various pan angles. The parameters of the pan-tilt model are the axis of pan motion, \(l_v\), expressed in the \(v\)-th camera coordinates; intersections of the axis of pan motion and the axis of tilt motion, \(c_{vi}\), expressed in the \(v\)-th camera coordinates; a pan angle \(\phi_{vi}\) of the \(i\)-th tilt rotation expressed in the \(v\)-th camera coordinates. Hereafter, the parameters are referred to as pan-tilt parameters, and \(c_{vi}\) is referred to as the pan-tilt center. The normal vector of the plane spanned by a tilt motion, \(n_{v,vi}\), is written by using \(l_v\) and \(\phi_{v,vi}\) easily.

Introducing the pan-tilt model makes it possible to express a rotation that transfers the original circle to circle \(C_{vi,vi}\), which is parallel to the image plane. This rotation can be expressed as

\[R_{PT} = R(l_v, \phi_{vi}) R\left(m_v, \sin^{-1} l_v/|l_v|\right) \quad (16)\]
where \( R \left( \mathbf{m}_v, \sin^{-1} l_{v3}/|l_v| \right) \) is a rotation around axis \( \mathbf{m}_v = l_v \times e_z \) by angle \( \sin^{-1} l_{v3}/|l_v| \). Here, \(|*|\) denotes the L2-norm. \( \mathbf{m}_v \) is an orthogonal vector to the axis of pan motion and parallel to the image plane, and \( e_z \) is a unit vector on the z-axis. This rotation transfers circle \( C_{c,v} \) to the one that is parallel to axis \( l_v \). By further rotation around the axis of pan motion, \( R(l_v, \phi_{v3}) \), the circle is moved to the original circle that corresponds to the observed trajectory.

According to the discussion in Sec. 4, rotation \( V_{vi} S_i R_{y,vi} R_z (-\alpha_{vi}) \) transfers circle \( C_{c,v} \) to the original circle. This rotation depends on observed ellipse \( Q_{vi} \) and an arbitrary parameter, \( \alpha_{vi} \). Unknown parameters of the pan-tilt model can therefore be estimated by minimizing the difference between the two rotations.

**Cost function for circle orientation:** Two types of constraints on circle orientations can be obtained as follows. The first can explain rotations fully. It includes rotation \( R_z (\alpha_{vi}) \). Since the solution for \( \alpha_{vi} \) using pan-tilt parameters is very complicated, \( \alpha_{vi} \) is used as an additional parameter. The second constraint is proposed to remove \( \alpha_{vi} \).

The first constraint, that is, a full-rotation constraint, is given as
\[
\| V_{vi} S_i R_{y,vi} - R(l_v, \phi_{v3}) R(m_v, \sin^{-1} l_{v3}/|l_v|) R_z(\alpha_{vi}) \| = 0 \quad (17)
\]
where \(|*|\) is the Frobenius norm. The equation means that the two rotations given by camera observation and the pan-tilt model are the same as those shown in Fig. 5. As for the second constraint, both sides of Eq. (17) are multiplied by the normal vector of circle \( C_{c,v} \) to obtain
\[
\left\{ V_{vi} S_i R_{y,vi} - R(l_v, \phi_{v3}) R(m_v, \sin^{-1} l_{v3}/|l_v|) \right\} n_{0,vi} = 0 \quad (18)
\]
where \( \phi_{v3} \) with various \( v \) denotes the direction of the normal vector of the same circle, it can be written with offset \( \theta_v \) as
\[
\phi_{v3} = \theta_v + \phi_k. \quad (20)
\]

Two types of cost functions, \( J_{rf} \) and \( J_{rn} \), that are based on squared errors of Eqs. (17) and (18), respectively, are thus obtained. Summations of the errors are thus given as
\[
J_{rf} (\alpha_{v1}, \ldots, \alpha_{v1}, \ldots, \alpha_{v1}, \ldots) = \sum_{v} \| V_{vi} S_i R_{y,vi} - R(l_v, \theta_v + \phi_v) R(m_v, \sin^{-1} l_{v3}/|l_v|) R_z(\alpha_{vi}) \|^2. \quad (21)
\]
\[
J_{rn} (\alpha_{v1}, \ldots, \theta_v, \ldots, \phi_v, \ldots, \alpha_{v1}, \ldots) = \sum_{v} \| V_{vi} S_i R_{y,vi} n_{0,vi} - R(l_v, \theta_v + \phi_v) R(m_v, \sin^{-1} l_{v3}/|l_v|) n_{0,vi,v} \|^2. \quad (22)
\]

**Cost function for circle center:** Three cost functions for circle centers are formulated as follows. The first cost function simply compares centers of the original circle and circle \( C_{c,v} \). The second cost function is obtained by removing additional parameter \( \alpha_{v} \) from the first one. The third one is introduced as a constraint that is independent of that on rotations. Minimizing it means to calculate the mean value of circle centers.

The first one uses the center of the original circle and circle \( C_{c,v} \). According to Fig. 5, the center of circle \( C_{c,v} \) is expressed using the pan-tilt model as
\[
c_{c,v} = (c_{c,v1} c_{c,v2} c_{c,v3})^T = R(m_v, -\sin^{-1} l_{v3}/|l_v|) R(l_v, -\phi_{v3}) c_v \quad (23)
\]

The first cost function is then obtained by comparing Eqs. (23) and (24) as follows
\[
J_{cf} (\alpha_{v1}, \ldots, \theta_1, \ldots, \phi_1, \ldots, \alpha_{v1}, \ldots) = \sum_{v} \| c_{c,v} - R_z (\alpha_{vi}) c_{c,v} \|^2. \quad (25)
\]

An additional parameter \( \alpha_{vi} \) from Eq. (25) is removed by comparing distances of the centers of the circles \( c_{c,v} \) and \( c_{x,v} \) from the origin, and the cost function is obtained as follows:
\[
J_{cm} (c_1, \ldots) = \sum_{v} \| V_{vi} S_i R_{y,vi} c_{x,v} - c_v \|^2. \quad (27)
\]

In summary, the two cost functions related to orientation, namely, \( J_{rf} \) and \( J_{rn} \), and the three cost functions related to position, namely, \( J_{rf} \), \( J_{rd} \) and \( J_{cm} \), are obtained as described above. Although \( J_{rf} \) and \( J_{rf} \) are complicated because of additional parameter \( \alpha_{vi} \), they fully express all rotations. The first cost function is given as \( J_{rf} + w J_{rf} \), where \( w \) is a weight. Since parameter \( \alpha_{vi} \) is removed, \( J_{rn} \), \( J_{rd} \), and \( J_{cm} \) have a simple form. \( J_{rf} + w J_{rd} \) is another possible cost function. Since \( J_{rn} \) and \( J_{cm} \) are independent, they...
can be respectively minimized. Thus, the following three cost functions are given as

\[
\text{Full method: } \min(J_{rf} + wJ_{ef})
\]
\[
\text{Alpha-free method: } \min(J_{rf} + wJ_{cd})
\]
\[
\text{Independent method: } \min J_{rn}; \min J_{em}
\]

Minimizing these cost functions gives \( l_v, \theta_v, \) and \( c_v \). Since these parameters are the same in the world coordinates, relative camera poses can be easily reconstructed.

6. Experiments

6.1. Experiments on simulated data

Setting: The proposed cost functions were evaluated by using simulated data. As shown in Fig. 1, it was assumed that a marker was set on a PTZ camera and trajectories of tilt motions could be observed as a sequence of tracked points. As shown in Fig. 6(A), two base pan angles, namely, -30 and +30 degrees, were set. According to the number of given tilt motions, other pan angles are equally spaced within \( \pm 5 \) degrees from the basic angles, as shown in Fig. 6(B). In these experiments, the number of cameras was set as two (since that is the most common expected situation). As for camera poses, four settings, as shown in Fig. 7, were adopted. Though camera pose (i), in which two cameras are at the same position, is not possible in a real situation, it was considered in this evaluation. As for other poses, poses (ii) is the one in which the directions of two target cameras are orthogonal, and poses (iii) and (iv) are those in which the target cameras and the reference camera are at different elevations. To simulate noise, Gaussian noise was added to each tracked marker point. To estimate ellipse parameters, hyper-renormalization [3] was used. All settings are summarized in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan angle</td>
<td>-25 to -35 and 25 to 35 (deg.)</td>
</tr>
<tr>
<td>Tilt angle</td>
<td>-10 to 70 (deg.)</td>
</tr>
<tr>
<td>Radius of tilt motion</td>
<td>55mm</td>
</tr>
<tr>
<td>Camera pose</td>
<td>4 types</td>
</tr>
<tr>
<td>Distance between</td>
<td>( d = 10.0 ) m</td>
</tr>
<tr>
<td>reference and target camera</td>
<td></td>
</tr>
<tr>
<td>Num. of sample points</td>
<td>100, 200, ..., 500 points per tilt motion</td>
</tr>
<tr>
<td>Noise of sample points</td>
<td>Gaussian noise, ( \sigma = 0.17, ) 0.33, 0.50, 0.67 (pix)</td>
</tr>
</tbody>
</table>

As for minimizing the cost functions, the Levenberg-Marquardt method was used. The constant \( w \) was set to 1. Initial values are given by the direct method with two circles observed by two cameras. Though the proposed method can handle the refinement with more than two cameras, the situation with two cameras only was evaluated because it is the most frequent and basic.

Experimental results: The direct method and the proposed optimization-based methods were evaluated. Single-view and multi-view refinement were also compared. The former minimized the cost functions by every \( \nu \)-th camera respectively, and the latter minimized them concerning all the cameras at the same time. To evaluate the accuracy of the methods, two types of error are defined as

\[
e_o = ||q - q_{GT}||, \quad e_p = ||t - t_{GT}||. \tag{28}
\]

Here, \( e_o \) is a quaternion error used to evaluate an error concerning orientation. \( q \) and \( q_{GT} \) are an estimated and a ground truth of quaternion that are converted from the rotation matrices of relative camera poses. \( e_p \) is a position error, where \( t \) and \( t_{GT} \) are estimated and ground-truth translations of the relative camera poses. All results shown in Figs. 8 to 15 are an average of 100 trials. In these experiments, the number of tracked points was 500, and the number of circles was 10. Amount of Gaussian noise was \( \sigma = 0.33 \) (pix) because it was assumed that amount of noise is almost within 1 pixel. In the figures, “D” stands for the direct method; “S-F,” “S-A,” and “S-I” stand for “single-view full method”, “single-view alpha-free method”, and “single-view independent method”; and, similarly, “M-F,” “M-A,” and “M-I” correspond to multi-view refinement.

Figs. 8 and 9 show that the alpha-free method is not good
at estimating camera position. A possible reason is that it has only two constraints concerning position estimation. The full method was not robust, especially in the case of multi-view optimization. Since the formulations of the full methods have one extra parameter, the optimization might become unstable. The best method was the independent method with multi-view optimization. As for the camera poses, since camera pose (i) is the simplest, it is reasonable that estimated camera pose was more accurate under that setting. Under camera poses (ii) and (iii), the observed circles were more oblique than those under the other camera poses. Since an oblique angle of observation deteriorated the estimation accuracy of ellipse fitting, the camera pose estimations were less accurate.

As for an effect of the number of tilt motions, the results are shown in Fig. 10 and 11. Hereafter, each result shown in figures is average of the result of four settings of camera poses. The number of tracked points was 500, and Gaussian noise was \( \sigma = 0.33 \) (pix). The optimization contributed to higher accuracy with a plural number of observed motions. When two or ten circles were assumed, the errors decreased by half. In this case, the independent method with multi-view optimization was the most accurate.

Fig. 12 and 13 show the effect of number of tracked points on camera pose estimation accuracy. The number of circles was 10, and Gaussian noise was \( \sigma = 0.33 \) (pix). Since a greater number of points made the ellipse fitting robust, accuracy of camera pose estimation was also improved. In other words, errors gradually decreased as number of tracked points increased. When the number was 500, the error was cut by a half. The independent method with multi-view optimization was again the most accurate.

Effect of noise is evaluated as Fig. 14 and 15. The number of circles was 10, and the number of tracked points was 500. It can be concluded from the figures that the error of the proposed method is linear in relation to noise.

**Comparison with point-wise method:** Since the marker is needed by the proposed method, point-wise estimation of camera pose is possible. However, due to network delays, keeping each pan-tilt state constant for a few seconds is required to assure temporal synchronizations between images captured by cameras. The number of feature points is limited in that case. This limitation is the same for the proposed (circle-wise) method, but it is only necessary to wait only once per tilt motion. Accordingly, as many feature points as possible can be captured during one motion. This is a major benefit of the proposed method.

To clarify the benefit of the proposed method, the point-wise method was also evaluated. Given 30 points that are three points on 10 circles respectively, the standard pipeline of 5pt algorithm, RANSAC, and bundle adjustment using Levenberg-Marquardt was performed. The quaternion error was very large under Gaussian noise of \( \sigma = 0.33 \) (pix). To obtain comparable estimation accuracies of 0.046, 0.040, 0.050, and 0.037 for camera poses (i)-(iv), small noise of \( \sigma = 0.033 \) (pix) had to be given.

**6.2. Experiments on real data**

The proposed method was applied to real cameras Canon VB-H41. Six rotations with pan angles of \( \pm 35, \pm 40, \) and \( \pm 45 \) degrees were used. First, intrinsic camera parameters were estimated using Zhang’s method [11] with a chessboard pattern. Relative pose was then estimated by using Zhang’s method and the proposed method.

As for the proposed method, input trajectories were captured by Camera 1 and Camera 2 as shown in Fig. 16. After normalization of the trajectories by using the intrinsic parameters, extracted points are used as input for the proposed method. As for the optimization, the independent method with multi-view optimization was selected.

Rotation matrix \( R \) and translation vector \( t \) (which translate a point in the camera coordinates of the Camera 2 to the one of the Camera 1) were estimated. By converting \( R \) into an expression composed of axis \( l \) and angle \( \theta \), the following result was obtained:

**Chessboard:**

\[ l = (0.025, -0.999, -0.031)^T \theta = 12.0(\text{deg}) \]
\[ t = (-810, -10, 212)^T (\text{mm}) \]

**Proposed method:**

\[ l = (0.012, -0.997, -0.081)^T \theta = 11.3(\text{deg}) \]
\[ t = (-818, -8, 241)^T (\text{mm}) \]

The difference between the axes estimated by the proposed method and the Zhang’s method was 2.93 degrees, and that between the angles was 0.7 degrees, and that between \( t \)s was \( (8 2 29)(\text{mm}) \). These results confirm that the proposed method can estimate a relative camera pose properly for the real cameras.

**7. Conclusion**

A novel method for estimating relative camera poses by using observed pan-tilt rotations is proposed. Since pan-tilt
motions are performed by the PTZ cameras themselves, the proposed method enables autonomous calibration of camera networks. By generating and observing pan-tilt motions each other, it is possible to determine relative poses between mutually visible cameras.

With the proposed method, circular trajectories generated by tilt motions with various pan angles are first observed. Since those trajectories are projected to the image plane as ellipses, ellipse fitting is performed. Normal vectors and centers of circles are then calculated. A novel refinement process to minimize the differences between modeled pan-tilt rotations and the observed ellipses is then performed. In particular, three types of cost functions were formulated. Experiments using simulated data confirmed the best cost function. Another experiment using real data showed that the proposed method could estimate the camera pose properly for real cameras.

As for future work, a system based on the proposed method for automatically calibrating a camera network will be developed. Especially, a method for detecting and identifying each camera in the network must be devised. A controllable PTZ function is convenient for this purpose because it can generate a specific pan-tilt motion for easy detection and identification of each camera.

References


