Calibration of an Articulated Camera System with Scale Factor Estimation

CHEN Junzhou*, Kin Hong WONG

Abstract

Multiple Camera Systems (MCS) have been widely used in many vision applications and attracted much attention recently. There are two principle types of MCS, one is the Rigid Multiple Camera System (RMCS); the other is the Articulated Camera System (ACS). In an RMCS, the relative poses (relative 3-D position and orientation) between the cameras are invariant. While, in an ACS, the cameras are articulated through movable joints, the relative pose between them may change. Therefore, through calibration of an ACS we want to find not only the relative poses between the cameras but also the positions of the joints in the ACS.

Although calibration methods for RMCS have been extensively developed during the past decades, the studies of ACS calibration are still rare. In this paper, we developed calibration algorithms for the ACS using a simple constraint: the joint is fixed relative to the cameras connected with it during the transformations of the ACS. When the transformations of the cameras in an ACS can be estimated relative to the same coordinate system, the positions of the joints in the ACS can be calculated by solving linear equations. However, in a non-overlapping view ACS, only the ego-transformations of the cameras and can be estimated. We proposed a two-steps method to deal with this problem. In both methods, the ACS is assumed to have performed general transformations in a static environment. The efficiency and robustness of the proposed methods are tested by simulation and real experiments. In the real experiment, the intrinsic and extrinsic parameters of the ACS are obtained simultaneously by our calibration procedure using the same image sequences, no extra data capturing step is required. The corresponding trajectory is recovered and illustrated using the calibration results of the ACS. Since the estimated translations of different cameras in an ACS may scaled by different scale factors, a scale factor estimation algorithm is also proposed. To our knowledge, we are the first to study the calibration of ACS.

I. INTRODUCTION

Calibration of a Multiple Camera System (MCS) is an essential step in many computer vision tasks such as SLAM (Simultaneous Localization and Map), surveillance, stereo and metrology [14], [3], [7], [9], [10], [17]. Both the intrinsic and extrinsic parameters of the MCS are required to be estimated before the MCS can be used. The intrinsic parameters [12], [11] describe the internal camera geometric and optical characteristics of each camera in the MCS. In a Rigid Multiple Camera System (RMCS), the cameras are fixed to each other. The extrinsic parameters [5] of a RMCS describe the relative pose (the relative 3-D position and orientation, totally, six degrees of freedom) between the cameras in the MCS. Calibration methods of the intrinsic parameters of a camera are well established [18], [21]. Calibration methods for the extrinsic parameters of a RMCS are also widely studied. For instance, in an RMCS, the positions of the joints in the ACS can be calculated relative to the cameras connected with it during the transformations of the ACS. When the transformations of the cameras in an ACS can be estimated relative to the same coordinate system, the positions of the joints in the ACS can be calculated by solving linear equations. However, in a non-overlapping view ACS, only the ego-transformations of the cameras and can be estimated. We proposed a two-steps method to deal with this problem. In both methods, the ACS is assumed to have performed general transformations in a static environment. The efficiency and robustness of the proposed methods are tested by simulation and real experiments. In the real experiment, the intrinsic and extrinsic parameters of the ACS are obtained simultaneously by our calibration procedure using the same image sequences, no extra data capturing step is required. The corresponding trajectory is recovered and illustrated using the calibration results of the ACS. Since the estimated translations of different cameras in an ACS may scaled by different scale factors, a scale factor estimation algorithm is also proposed. To our knowledge, we are the first to study the calibration of ACS.

Scale Factor Estimation
calibration methods cannot estimate the positions of the joints in the ACS. (ii) In a non-overlapping view ACS, neither the positions of the joints in the ACS nor the relative poses between the cameras in the ACS can be estimated by traditional calibration methods.

These considerations in mind motivate us to develop the technologies in this paper. The rest of this paper are organized as follows: Section II and III analysis the constraints in a moving ACS. The corresponding calibration methods are proposed. Section IV and V evaluate the proposed method by simulation and real experiment. In section VI, a brief conclusion and the future plan are presented.

II. CALIBRATION OF ACS WITH OVERLAPPING VIEWS

Suppose two rigid objects are articulated at joint O and two cameras (camera A and B) are fixed on the two rigid objects respectively (See Figure 2). Let $C_A$ be the coordinate system of camera A, $C_B$ the coordinate system of camera B. Suppose there are enough feature correspondences between the cameras so that the pose of $C_A$ and $C_B$ referring to the same coordinate system $C_W$ can be estimated. Therefore, the relative pose between $C_A$ and $C_B$ is known. We want to find the position of O in the ACS. Let $H_{AW}$ and $H_{BW}$ be the Euclidean transformation matrixes describe the $C_A$ and $C_B$ relative to $C_W$, so that for any point $P$:

\[
P_A = H_{AW}P_W = \begin{bmatrix} \mathbf{R}_{AW} & T_{AW} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{P}_W \\ 1 \end{bmatrix}
\]

\[
P_B = H_{BW}P_W = \begin{bmatrix} \mathbf{R}_{BW} & T_{BW} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{P}_W \\ 1 \end{bmatrix}
\]

, where $\mathbf{R}$ is the $3 \times 3$ rotation matrix, $T$ is a $3 \times 1$ vector, $P_W$, $P_A$ and $P_B$ are the homogenous coordinates of the 3-D Point $P$ relative to $C_W$, $C_A$ and $C_B$ respectively, $\bar{P}$ is a $3 \times 1$ vector.
According to equations (1) and 2:

$$P_W = H^{-1}_{AW} P_A = H^{-1}_{BW} P_B$$

$$H^{-1}_{AW} P_A - H^{-1}_{BW} P_B = 0$$

$$\begin{bmatrix} R^T_{AW} & -R^T_{AW} T_{AW} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_A \\ 1 \end{bmatrix} - \begin{bmatrix} R^T_{BW} & -R^T_{BW} T_{BW} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_B \\ 1 \end{bmatrix} = 0$$

$$R^T_{AW} P_A - R^T_{BW} P_B = R^T_{AW} T_{AW} - R^T_{BW} T_{BW}$$

where $R^T$ is the transpose of $R$. Suppose the ACS performed $n$ transformations. Let $H^i_{AW}$ and $H^i_{BW}$ be the Euclidean transformation matrixes describe the $C_A$ and $C_B$ relative to $C_W$ after the $i$-th transformation of the ACS. According to equation (6):

$$(R^i_{AW})^T P_A - (R^i_{BW})^T P_B = (R^i_{AW})^T T_{AW}^i - (R^i_{BW})^T T_{BW}^i$$

Let $\tilde{O} = [\tilde{O}_A^T \tilde{O}_B^T]^T$, where $\tilde{O}_A$ and $\tilde{O}_B$ are the coordinates of the joint $O$ relative to $C_A$ and $C_B$ respectively. Equation (7) can be rewritten as:

$$\begin{bmatrix} (R^i_{AW})^T \\ -(R^i_{BW})^T \end{bmatrix} \tilde{O} = (R^i_{AW})^T T_{AW}^i - (R^i_{BW})^T T_{BW}^i$$

Since camera A and B are fixed on the articulated rigid objects, $\tilde{O}$ is invariant during the transformation of the ACS. The transformations $(R^i_{AW}, R^i_{BW}, T_{AW}^i$ and $T_{BW}^i$ for $i \in [1 \ldots n]$) of the camera coordinate systems are calculated by the projected image sequences. We propose that $\tilde{O}$ can be estimated by a least squares method, when the ACS has moved to many different positions and captured enough samples of $R^i_{AW}, R^i_{BW}, T_{AW}^i$ and $T_{BW}^i$.

The above derivation shows that although the location of the joint $O^i_W$ in world coordinates is not constant, it equals $(H^i_{AW})^{-1} O_A$ or $(H^i_{BW})^{-1} O_B$ because the cameras can not move completely independent as they are connected with a joint. The joint location can be calculated by the 1D subspace intersection of the camera transformation matrices.

### III. CALIBRATION OF NON-OVERLAPPING VIEW ACS

In many situations, there is no overlapping view between the cameras in an ACS. And the lack of common features makes the calibration method proposed in section II become invalid (See Figure 3). Moreover, since the relative pose between the cameras in the ACS cannot be estimated by the overlapping views, the calibration of the relative poses between the non-overlapping view cameras is also required. In this section, a calibration method based on the ego-motion information of the cameras in an ACS is discussed.

#### A. Recovering the Position of the Joint Relative to the Cameras in the ACS

Let $C_A^{init}$ and $C_B^{init}$ be the coordinate systems of camera A and B respectively at the initial state (time $t = 0$). Suppose the ACS performs $n$ transformations. Since the coordinate of the joint O relative to camera A is fixed during the transformation of the ACS. At time $t = i$, we have:

$$O^i_A = H^i_A O_A = \begin{bmatrix} R^i_A & T^i_A \\ 0 & 1 \end{bmatrix} O_A$$

$$\begin{bmatrix} R^i_A & T^i_A \\ 0 & 1 \end{bmatrix}$$
where \( H_A^i \) is the Euclidean transformation matrix of camera A at time \( i \) relative to \( C_A^{\text{init}} \). \( R_A^i \) and \( T_A^i \) describe the orientation and origin of camera A at time \( i \) relative to \( C_A^{\text{init}} \). Also \( O_A \) is the coordinate of point O at initial state relative to \( C_A^{\text{init}} \), and \( O_A^i \) is the coordinate of point O at time \( i \) relative to \( C_A^{\text{init}} \).

If the position of the joint O relative to \( C_A^{\text{init}} \) is fixed during the transformations of the ACS, we have: \( O_A^i = O_A, \forall i \in [1, \ldots, n] \). For \( i \)-th transformation of the ACS, according to equation (9):

\[
O_A = H_A^i O_A = \begin{bmatrix} R_A^i & T_A^i \\ 0 & 1 \end{bmatrix} O_A
\]

(10)

\[
(R_A^i - I) \hat{O}_A = - T_A^i
\]

(11)

Let \( M_A = [(R_A^1 - I)^T, (R_A^2 - I)^T, \ldots, (R_A^n - I)^T]^T, \hat{T}_A = [(T_A^1)^T, (T_A^2)^T, \ldots, (T_A^n)^T]^T \), we have:

\[
M_A \hat{O}_A = - \hat{T}_A
\]

(12)

Since the transformations \( (R_A^i, T_A^i, \forall i \in [1 \ldots n]) \) of camera A can be calculated by the projected image sequence. We propose \( \hat{O}_A \) can be estimated by a least squares method. Similarly, \( \hat{O}_B \) can also be estimated. Therefore, \( O_A \) and \( O_B \) are recovered.

B. The Uniqueness of the Joint Pose Estimation

If the different segments of the articulated camera system (ACS) are connected by 1D rotational joints (connected by point rotational joints) and the ACS can perform general transformations, the solution of the joint pose estimation is unique:

For the joint pose estimation method using special motion (in section III-A). Suppose the solution of the joint pose estimation is not unique, there must exist at least two different 3D points \( \hat{O}_1 \) and \( \hat{O}_2 \) satisfy equation (12). We have: \( M_A \hat{O}_1 = - \hat{T}_A \) and \( M_A \hat{O}_2 = - \hat{T}_A \). Therefore, any point \( \hat{P} = s \hat{O}_1 + (1 - s) \hat{O}_2 \) will also satisfy equation (12), where s is an arbitrary scalar. According to the definition of \( \hat{P}, \hat{P} \) is the point on the line passing through the points \( \hat{O}_1 \) and \( \hat{O}_2 \). Since \( \hat{P} \) satisfy equation (12) represents that the position of the point P relative to the camera in the ACS is invariant during the transformation of the ACS, it means the different segments of ACS are connected by the 2D rotational axis instead of the 1D rotational joints. The position of the points on the 2D rotational axis relative to the camera in the ACS is invariant during the transformation of the ACS. However, it conflicts with the assumption. Similarly, the uniqueness of the joint pose estimation method using overlapping views (in section III) can also be verified.

C. Recovering the Relative Pose Between the Cameras of the Non-overlapping view ACS

Let \( H_{BA} \) be the Euclidean transformation matrix between \( C_A^{\text{init}} \) and \( C_B^{\text{init}} \), so that for any point \( P \):

\[
P_B = H_{BA} P_A = \begin{bmatrix} R_{BA} & T_{BA} \\ 0 & 1 \end{bmatrix} P_A = H_{BA} P_A
\]

(13)

, where \( P_A \) and \( P_B \) are the homogenous coordinate of Point P relative to \( C_A^{\text{init}} \) and \( C_B^{\text{init}} \) respectively.

The relative pose (\( \hat{R}_{BA} \) and \( \hat{T}_{BA} \)) between \( C_A^{\text{init}} \) and \( C_B^{\text{init}} \) is defined as:

\[
\hat{R}_{BA} = R_{BA}^T
\]

(14)

\[
\hat{T}_{BA} = - R_{BA}^T \hat{T}_{BA}
\]

(15)

Let \( O_B^i \) be the coordinate of joint O at time \( i \) relative to \( C_B^{\text{init}} \). Since the coordinate of the joint O relative to camera B is invariant:

\[
O_B^i = \begin{bmatrix} R_B^i & T_B^i \\ 0 & 1 \end{bmatrix} O_B
\]

\[
= \begin{bmatrix} R_B^i & T_B^i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{BA} & T_{BA} \\ 0 & 1 \end{bmatrix} O_A
\]

\[
= \begin{bmatrix} R_B R_{BA} & R_B \hat{T}_{BA} + T_B \\ 0 & 1 \end{bmatrix} O_A
\]

(16)

According to equations (9) and (13):

\[
O_B^i = H_{BA} O_A^i
\]

\[
\begin{bmatrix} R_{BA} & T_{BA} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_A^i & T_A^i \\ 0 & 1 \end{bmatrix} O_A
\]

\[
= \begin{bmatrix} R_B R_{BA} & R_B \hat{T}_{BA} + T_B \\ 0 & 1 \end{bmatrix} O_A
\]

(17)
According to equations (16) and (17):

\[
\begin{bmatrix}
R_B^i R_{BA} & R_B^i T_{BA} + T_B^i \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
O_A \\
1
\end{bmatrix}
= \begin{bmatrix}
R_{BA}^i R_A^i & R_{BA}^i T_A^i + T_{BA} \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
O_A \\
1
\end{bmatrix}
\]

(18)

\[
\begin{bmatrix}
R_B^i R_{BA} \bar{O}_A + R_B^i T_{BA} + T_B^i \\
0 & 1
\end{bmatrix}
= \begin{bmatrix}
R_{BA}^i R_A^i \bar{O}_A + R_{BA}^i T_A^i + T_{BA} \\
0 & 1
\end{bmatrix}
\]

(19)

\[
R_B^i R_{BA} \bar{O}_A + R_B^i T_{BA} - R_{BA}^i R_A^i \bar{O}_A - R_{BA}^i T_A^i + T_B^i - T_{BA} = 0
\]

(20)

Since \( \bar{O}_A \) can be estimated by the method discussed in section III-C, the \( R_{BA} \) and \( T_{BA} \) can be estimated by a least square method, when the ACS perform enough general motions.

In our simulation and real experiment, the estimated \( R_{BA} \) is refined by a method discussed in [20]. Then the roll, pitch and yaw corresponding to the \( R_{BA} \) are estimated according to the definition of the rotation matrix [11]. Let \( R_{BA} = M(r, p, y) \), where \( r \) \( p \) and \( y \) are the corresponding roll, pitch and yaw of \( R_{BA} \). \( M \) is a function from roll, pitch and yaw to the corresponding rotation matrix. Then, the \( r \), \( p \), \( y \), \( T_A \) and \( \bar{O}_A \) are estimated by minimizing the nonlinear error function:

\[
E(r, p, y, T_A, \bar{O}_A) = \sum_{i=1}^{n} (R_B^i M(r, p, y)\bar{O}_A + R_B^i T_{BA})
\]

\[
- M(r, p, y)R_A^i \bar{O}_A - M(r, p, y)T_A^i + T_B^i - T_{BA}
\]

(21)

using a Levenberg-Marquardt method. Finally, the \( R_{BA} \) is recovered from the optimized \( r \), \( p \) and \( y \). The relative pose between the \( C_A^{init} \) and \( C_B^{init} \) is calculated by equations (14) and (15).

IV. DEALING WITH UNKNOWN SCALE FACTORS

The non-overlapping view ACS calibration method discussed above depends on the ego-motion information of the cameras in the ACS. However, if the model of the scene is unknown, the estimated ego-translations of the cameras may be scaled by different unknown scale factors. These unknown scale factors must be considered in the extrinsic calibration process.

A. Model Analysis

Let \( T_A \) and \( T_B \) be the true ego-translation of camera A and B in the world coordinate system, \( \hat{T}_A \) and \( \hat{T}_B \) be the estimated ego-translations of camera A and B found by an SFM method, \( \mu_A \) and \( \mu_B \) be the corresponding unknown scale factors. So that:

\[
\hat{T}_A = \mu_A T_A
\]

(22)

\[
\hat{T}_B = \mu_B T_B
\]

(23)

Let \( \hat{O}_A \) be the pose of the joint relative to \( C_A \) calculated with the estimated motion. Equation (11) can be rewritten as:

\[
(R_A^i - I)\hat{O}_A = -\hat{T}_A^i = -\mu_A T_A^i
\]

(24)

\[
(R_A^i - I) \frac{\hat{O}_A}{\mu_A} = -T_A^i
\]

(25)

Compare equation (25) with equation (11), we have:

\[
\hat{O}_A = \mu_A \bar{O}_A
\]

(26)

Let \( \hat{R}_{BA} \) and \( \hat{T}_{BA} \) be the extrinsic parameters calculated using the estimated motions and joint pose. Equation (20) can be rewritten as:

\[
R_B^i \hat{R}_{BA} \hat{O}_A + R_B^i \hat{T}_{BA} - \hat{R}_{BA} R_A^i \bar{O}_A - \hat{R}_{BA} T_A^i + \hat{T}_B^i - \hat{T}_{BA} = 0
\]

(27)

According to equation (22), (23), (26) and (27):

\[
R_B^i \mu_A \hat{R}_{BA} \bar{O}_A + R_B^i \hat{T}_{BA} - \hat{R}_{BA} R_A^i \mu_A \bar{O}_A - \hat{R}_{BA} \mu_A T_A^i + \mu_B T_B^i - \hat{T}_{BA} = 0
\]

(28)

\[
R_B^i \frac{\mu_A}{\mu_B} \hat{R}_{BA} \bar{O}_A + R_B^i \frac{1}{\mu_B} \hat{T}_{BA} - \frac{\mu_A}{\mu_B} \hat{R}_{BA} R_A^i \bar{O}_A - \frac{\mu_A}{\mu_B} \hat{R}_{BA} T_A^i + \frac{1}{\mu_B} \hat{T}_{BA} = 0
\]

(29)

Let:

\[
\frac{\mu_A}{\mu_B} \hat{R}_{BA} = \hat{R}_{BA}
\]

(30)
\[
\frac{1}{\mu_B} \dot{T}_{BA} = \ddot{T}_{BA}
\]  
(31)

Equation (29) can be rewritten as:
\[
R_B^i R_{BA} \dot{O}_A + R_B^i \ddot{T}_{BA} - \dot{R}_{BA} R_A^i \dot{O}_A - \dot{R}_{BA} T_A^i + T_B - \ddot{T}_{BA} = 0
\]  
(32)

Since the equations (32) and (20) are exactly the same, we have:
\[
R_{BA} = \dot{R}_{BA}
\]  
(33)
\[
T_{BA} = \dddot{T}_{BA}
\]  
(34)

Therefore:
\[
\dot{R}_{BA} = \frac{\mu_B}{\mu_A} R_{BA} = \phi_{BA} R_{BA}
\]  
(35)
\[
\dddot{T}_{BA} = \mu_B \dddot{T}_{BA} = \mu_B T_{BA}
\]  
(36)

Where \( \phi_{BA} = \frac{\mu_B}{\mu_A} \). Equations (35) and (36) show that the estimated rotation matrix \( \dot{R}_{BA} \) will be scaled by the relative scale factor (the ratio of the scale factors of the cameras) and the estimated relative translation will be scaled by the same scale factor of camera \( B \). In the next section, we will discuss the estimation of the relative scale factor.

**B. Rotation Matrix and Relative Scale Factor Estimation**

Let \( R' = \phi R + N \), where \( R \) is a \( 3 \times 3 \) rotation matrix and \( R^T R = I \), \( \phi \) is an unknown scale factor, \( N \) is a \( 3 \times 3 \) unknown noise matrix. We want to recover \( R \) and \( \phi \) from \( R' \). According to the definition, we have:
\[
R' = \phi R + N = \phi(R + \frac{N}{\phi}) = \phi M
\]  
(37)

Where \( M = R + \frac{N}{\phi} \).

Let the singular value decomposition of \( M \) be \( U D V^T \), where \( D = \text{diag}(\sigma_1, \sigma_2, \sigma_3) \). As illustrated in appendix C of \( [20] \), \( r \) can be approximated by:
\[
R = U V^T
\]  
(38)

Now, let the singular value decomposition of \( r' \) be \( \tilde{U} \tilde{D} \tilde{V}^T \), since \( R' = \phi M \), we have:
\[
\tilde{U} = U
\]  
(39)
\[
\tilde{V} = V
\]  
(40)
\[
\tilde{D} = \phi D
\]  
(41)

Combine equations (38), (39) and (40), the rotation matrix \( r \) can be recovered by:
\[
R = \tilde{U} \tilde{V}^T
\]  
(42)

When noise \( N \) is not significant, \( D \approx I_{3 \times 3} \), the scale factor \( \phi \) can be estimated by the following approximation:
\[
\text{trace}(\tilde{D}) = \text{trace}(\phi D) \approx \text{trace}(\phi I_{3 \times 3}) \approx 3 \phi
\]  
(43)
\[
\phi \approx \frac{1}{3} \text{trace}(\tilde{D})
\]  
(44)

In short, if we have enough samples of \( R_{BA}^i, T_{BA}^i \), \( R_B^i \) and \( T_B^i \) we can find \( \dot{O}_A, \dot{R}_{BA} \) and \( \dddot{T}_{BA} \) (see section IV-A). Then using the above formulas, in particular, equation (42) and (44), we can also find the real rotation (\( R_{BA} \)) and the relative scale factor \( \phi_{BA} \).

Let \( R_{BA} = M(r, p, y) \), where \( r, p \) and \( y \) are the corresponding roll, pitch and yaw of \( R_{BA} \). \( M \) is a function from roll, pitch and yaw to the corresponding rotation matrix. In our simulation and real experiment, the estimated \( r, p, \) and \( y, \dddot{T}_{BA} \) can be optimized by minimizing the nonlinear error function:
\[
E(r, p, y, T_{BA}, Q_A) = \sum_{i=1}^{N} (\phi_{BA} R_B^i M(r, p, y) \dot{O}_A + R_B^i \dddot{T}_{BA} - \phi_{BA} M(r, p, y) R_A^i \dot{O}_A - \phi_{BA} M(r, p, y) T_A^i + T_B - \dddot{T}_{BA})
\]  
(45)

using a Levenberg-Marquardt method. If the pose of the joint is calibrated with known scale factor (\( O_A \) is known), the scale factor \( \mu_A \) can be estimated by equation (26). The scale factor \( \mu_B \) can be calculated by \( \frac{\phi_{BA}}{\phi_{BA}} \). Finally, the \( R_{BA} \) is recovered from the optimized \( r, p \) and \( y \). The relative pose between the \( C_{A}^{init} \) and \( C_{B}^{init} \) is calculated by equations (14) and (15). Therefore, a non-overlapping view ACS can also be calibrated using scaled motion information from each camera in it.
V. SIMULATION

In this section, the proposed calibration methods are evaluated with synthetic transformation data.

A. Performance w.r.t. Noise in Transformation Data

Setup and Notations: In each test, one ACS with 2 cameras and 1 joint is generated randomly. In which, \(1 \leq |O_A| \leq 2\) meters, \(1 \leq |O_B| \leq 2\) meters. The generated ACS performs 30 random transformations.

Performance of the Calibration Method for ACS with Overlapping Views: In the first simulation, the proposed algorithm is tested 100 times. Zero mean Gaussian noise is added to the transformation data of the cameras. The configuration, input and output of our simulation system are list as Table I. Since we assume there are overlapping views between the two cameras, the relative pose between them can be estimated by many existing methods as discussed in section I. Only the performance of joint pose estimation is evaluated in our simulation. The error of joint estimation are computed by:

\[
Err = \frac{|O_A - \hat{O}_A|}{2|O_A|} + \frac{|O_B - \hat{O}_B|}{2|O_B|} \tag{46}
\]

, where \(O_A\) is the ground truth, \(\hat{O}_A\) is the estimated position of joint O relative to camera A. Similarly, \(O_B\) is the ground truth, \(\hat{O}_B\) is the estimated position of joint O relative to camera B. The corresponding results are shown in Figure 4.

\[
\begin{array}{|c|c|c|}
\hline
\text{Configuration} & \text{Input (} i = 1 \ldots n \text{)} & \text{Output} \\
\hline
\text{No. of Cameras in the ACS} & 2 & \text{Mean error of joint pose estimation (see equation (46))} \\
\text{No. of Joints in the ACS} & 1 & \text{STD error of joint pose estimation (see equation (46))} \\
\text{Random transformations per test (n)} & 30 & \\
\text{Number of tests} & 100 & \\
\hline
\text{Rotations of cameras (} R_{AW}^i, R_{BW}^i \text{)} & 2 \times 30 \times 100 & \text{Translation Noise (meter)} \\
\text{Translations of cameras (} T_{AW}^i, T_{BW}^i \text{)} & 2 \times 30 \times 100 & \\
\text{Zero Mean Gaussian noise!} & 0 \leq \sigma_{\text{rot}} \leq 2.4^\circ \text{ and } 0 \leq \sigma_{\text{trans}} \leq 0.1 \text{meters} & \\
\hline
\end{array}
\]

Performance of the Calibration Method for Non-Overlapping Views ACS: In the second simulation, firstly, the pose of the joint is fixed relative to \(C_A^{\text{init}}\) during the transformations of the ACS. The pose of the joint relative to the camera A (\(O_A\)) is calibrated by the transformations of camera A. Similarly, \(O_B\) is calibrated. Then, the ACS performs several general transformations (the joint is not needed to be fixed relative to \(C_A^{\text{init}}\)), the relative pose between the cameras are calibrated using the estimated joint pose and the transformations of the cameras. The configuration, input and output of the simulation system are listed as Table II. The error of joint pose, relative rotation, relative translation estimation are calculated by equation (46), (47) and (48) respectively.

![Fig. 4](image-url)

**Fig. 4**

**Mean and STD Error of Joint Pose \((O_A)\) Estimation.** (a) **Mean Error of Joint Pose Estimation;** (b) **STD Error of Joint Pose.**
Figure 5 shows the results of joint pose estimation. Compare with the calibration method using the overlapping views, the calibration method using special motions is more accurate. The mean and STD error of the relative rotation and translation estimation are presented in Figure 6 and 7. The proposed algorithms are shown to be stable, when the zero mean Gaussian noise from $0^\circ$ to $2.4^\circ$ is added to the roll, pitch and yaw of the rotation data, and the zero mean Gaussian noise from 0 to 0.1 meters is added to the translation data.

\[
E_{err}^{rot} = \sqrt{|roll - \hat{roll}|^2 + |pitch - \hat{pitch}|^2 + |yaw - \hat{yaw}|^2} \tag{47}
\]

\[
E_{err}^{trans} = \frac{|T_{AB} - \hat{T}_{AB}|}{|T_{AB}|} \tag{48}
\]

### TABLE II
**CONFIGURATION, INPUT AND OUTPUT**

| Configuration     | No. of Cameras in the ACS | No. of Joints in the ACS | Random transformations per test (n) | Number of tests |
|-------------------|---------------------------|--------------------------|------------------------------------|-----------------|
| Input (i = 1 . . . n) |                           |                          |                                    |                 |
| Transforms with fixed joint pose: |                           |                          |                                    |                 |
| Rotations of cameras (R_{iA}, R_{iB}) | 2 × 30 × 100     |                         |                                    |                 |
| Translations of cameras (T_{iA}, T_{iB}) | 2 × 30 × 100      |                         |                                    |                 |
| General transforms: |                           |                          |                                    |                 |
| Rotations of cameras (R_{iA}, R_{iB}) | 2 × 30 × 100     |                         |                                    |                 |
| Translations of cameras (T_{iA}, T_{iB}) | 2 × 30 × 100      |                         |                                    |                 |
| Zero Mean Gaussian noise: |                           |                          |                                    |                 |
| 0 ≤ σ_{rot} ≤ 2.4° and 0 ≤ σ_{trans} ≤ 0.1meters |                           |                          |                                    |                 |

| Output                                   | Mean error of joint pose estimation (see equation (46)) | STD error of joint pose estimation (see equation (46)) | Mean error of relative translation estimation (see equation (48)) | STD error of relative translation estimation (see equation (48)) | Mean error of relative rotation estimation (see equation (47)) | STD error of relative rotation estimation (see equation (47)) |

![Fig. 5](image)

**Mean and STD Error of Joint Pose (\(\hat{O}_A\)) Estimation.** (a) Mean Error of Joint Pose Estimation; (b) STD Error of Joint Pose Estimation.

**Performance of the Calibration Method for Non-Overlapping Views ACS with Unknown Scale Factors:** The scale factors of the two cameras in each test are assumed to be uniform distributed in the range [0.5, 5]. Therefore, the relative scale factor between the two cameras satisfies the uniform distribution in the range of [0.1, 10]. The joint pose of the ACS is generate randomly and estimated by the method described in section III-A. Other configurations are the same as the second simulation. The \(\hat{O}_A\), \(R_{BA}\), \(T_{BA}\) and \(\phi_{BA}\) are estimated and optimized as discussed in section IV. The error of joint pose, relative rotation, relative translation estimation are calculated by equation (46), (47) and (48) respectively. The error of relative...
scale factor estimation is evaluated by $\varepsilon_\phi = \frac{|\phi - \hat{\phi}|}{|\phi|}$. Where $\hat{\phi}$ is the estimated relative scale factor, and $\phi$ is the ground truth.

Figure 9 and 10 show the results of the relative pose estimation. Compared to figure 6 and 7 the accuracies are similar. Figure 11 shows the performance of the relative scale factor estimation. The accuracy of the relative scale factor estimation $[(1 - \varepsilon_\phi) \times 100\%]$ is no less than 98.5%, when the standard derivation of the noise in ego-rotation is less than 3° and the standard derivation of the noise in ego-translation is less than 0.1 meters.

VI. REAL EXPERIMENT

In the real experiments, an ACS with two cameras (Cannon PowerShot G9) is set up as Figure 13(a). The intrinsic parameters of each camera in the ACS are calibrated by Bouguet’s implementation (“Camera Calibration Toolbox for Matlab”) of [21]. Since the Bouguet’s Toolbox can also estimate the pose information of the camera, the transformations of each camera are calculated using the same image sequence for the intrinsic calibration simultaneously. No additional images nor manual input is required in the real experiments.

A. Calibration of the Pose of the Joint in Each Camera

By Overlapping Views (Algorithm I): In the first real experiment, the two cameras in the ACS observe the same checker plane and record images simultaneously. The two cameras are free to move during the transformation of the ACS. Two image
sequences ($Q_1$ and $Q_2$) are recorded, each sequence consists of 15 images of size 1600 × 1200 pixels. The estimated joint pose are list in Table III as algorithm I.

By Fixed-Joint Motions (Algorithm II): In the second real experiment, the joint of the ACS is fixed relative to the world coordinate system during the transformation of the ACS. The two cameras do not need to view the same checker plane. And each camera records the image sequence independently. Two image sequences ($Q_3$ and $Q_4$) are recorded, each sequence consists of 12 images of size 1600 × 1200 pixels. The camera pose of the first image is selected as the initial pose to generate the transformation sequence of each camera. The estimated joint pose are list in Table III as algorithm II. The poses of the joint relative to the two cameras in the ACS are also estimated manually for comparison purpose. Since the camera pose of any image in each image sequence can be chosen as the initial camera pose (see section III-A), the proposed algorithm is also tested by choosing different images as the reference. The mean and standard derivation of the corresponding calibration results are presented in Table IV.

B. Calibration of Relative Pose Between the Cameras in the Non-Overlapping View ACS (Algorithm III)

In the third real experiment, firstly, we use the non-overlapping view ACS calibration method to process the image sequences $Q_1$ and $Q_2$. The joint pose ($\hat{O}_A$) estimated by algorithm II is used as the input for the relative pose calibration. Since there are overlapping views between $Q_1$ and $Q_2$, we also calibrate the relative pose between the two cameras by the feature correspondences for comparison. The calibration result are listed in Table V. After the joint pose relative to each camera in the ACS and relative pose between the cameras in the ACS are calibrated, the trajectory of the ACS is recovered (see Figure 12). The proposed calibration method is also tested by non-overlapping view image sequences. Figure 13 (b), (c), (d) shows the configuration of the non-overlapping view ACS calibration system in the real experiment. Two image sequences ($Q_5$ and
Q6) are recorded, each sequence consists of 17 images of size 1600 × 1200 pixels. There is no overlapping view between Q5 and Q6. Figure 14 shows some samples of the recorded images. We also manually measured the relative pose between the two cameras for comparison. Since no feature correspondence can be used, we only get a rough estimation by a ruler. The calibration results are shown in Table VI. After the relative pose between the cameras at the initial state is estimated, the trajectory of the non-overlapping view ACS is recovered (see Figure 15).

C. Calibration of Relative Pose Between the Cameras in the Non-Overlapping View ACS with Unknown Scale Factors (Algorithm IV)

The scale factor estimation algorithm is evaluated in the fourth real experiment. The estimated translations from Q1 and Q2 are multiplied by 0.8 and 3.2 respectively. In this case, if no noise exists, the estimated relative scale factor (φBA) should be 4. The estimated relative scale factor (φBA) in our experiment was 3.8919. Table VII lists the corresponding results, in which the estimated relative translations are divided by 3.2, so that they can be easily compared with the estimated relative translations in Table VI. The experiment showed that our algorithms can estimate the relative scale factor and find the extrinsic parameters correctly. In order to test the stability of the scale factor estimation algorithm, the estimated translations from Q5 and Q6 are multiplied by 0.8 and 3.2 respectively. 100 tests are performed. In each test, 22 images are randomly selected as section VI-B. The Mean and STD of the calibration results is listed in Table VIII. The results are good.

VII. CONCLUSION

In this paper, an ACS calibration method is developed. Both the simulation and real experiment show that the pose of the joint in an ACS can be estimated robustly. When there is no overlapping view between the cameras in an ACS, the joint pose
TABLE III
RESULTS OF JOINT POSE CALIBRATION

I: the algorithm using overlapping views. (see section VI-A) II: the algorithm using fixed-joint motions. (see section VI-A) M: manual measurement (ground truth). $O_A$ is the coordinate of the joint relative to camera $A$, the same applies to $O_B$.

| Algorithm | Joint Pose (mm) |
|-----------|-----------------|
|           | $A$             | $Y$           | $Z$           |
| I         | $O_A$           | 380.28        | 50.04         | -33.47       |
|           | $O_B$           | -223.70       | 53.81         | -30.15       |
| II        | $O_A$           | 304.55        | 47.64         | -37.66       |
|           | $O_B$           | -265          | 54.41         | -35.48       |
| M         | $O_A$           | 300 ± 10      | 50 ± 10       | -40 ± 10     |
|           | $O_B$           | -270 ± 10     | 50 ± 10       | -30 ± 10     |

TABLE IV
MEAN AND STD OF THE JOINT POSE CALIBRATION ALGORITHM II USING DIFFERENT REFERENCE IMAGES

$O_A$ is the coordinate of the joint relative to camera $A$, the same applies to $O_B$.

| Algorithm | Joint Pose (mm) |
|-----------|-----------------|
|           | $X$             | $Y$           | $Z$           |
| Mean      | $O_A$           | 305.44        | 47.19         | -39.2        |
|           | $O_B$           | -262.97       | 56.21         | -39.20       |
| STD       | $O_A$           | 1.89          | 1.16          | 3.02         |
|           | $O_B$           | 3.3           | 2.67          | 2.58         |

and the relative pose between the cameras can also be calculated. The trajectory of an ACS can be recovered after the ACS is calibrated. The proposed calibration method requires only the image sequences recorded by the cameras in the ACS. A scale factor estimation algorithm is proposed to deal with unknown scale factors in the estimated translation information of the cameras in an ACS. In the real experiment, the intrinsic and extrinsic parameters of the ACS are calibrated using the same image sequences simultaneously.

Since we still cannot find any former study of the ACS calibration in the literature. We apologize for having no comparison with former ACS calibration method.

Our future plan may focus on using an ACS attached on different parts of human body to track the motion of the human. We foresee that if calibration of articulated cameras become a simple routine, researchers will find many novel and interesting applications for such a camera system.

REFERENCES

[1] M. Antone and S. Teller. Scalable extrinsic calibration of omni-directional image networks. International Journal of Computer Vision, 49(2):143–174, 2002.
[2] P. Baker and Y. Aloimonos. Complete calibration of a multi-camera network. Proc. IEEE Workshop on Omnidirectional Vision, 12:134–141, 2000.
[3] P. Baker, A. Ogale, and C. Fermuller. The Argus eye: a new imaging system designed to facilitate robotic tasks of motion. Robotics & Automation Magazine, IEEE, 11(4):31–38, 2004.
[4] P. T. Baker and Y. Aloimonos. Calibration of a multicamera network. Conference on Computer Vision and Pattern Recognition Workshop, 07, 2003.
[5] B. Caprile and V. Torre. Using vanishing points for camera calibration. International Journal of Computer Vision, 4(2):127–139, 1990.
[6] Y. Caspi and M. Irani. Aligning Non-Overlapping Sequences. International Journal of Computer Vision, 48(1):39–51, 2002.
[7] S. Dockstader and A. Tekalp. Multiple camera tracking of interacting and occluded human motion. Proceedings of the IEEE, 89(10):1441–1455, 2001.
[8] F. Dornaika. Self-calibration of a stereo rig using monocular epipolar geometries. Pattern Recognition, 40(10):2716–2729, 2007.
[9] Y. Furukawa and J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision, 84(3):257–268, 2009.
[10] R. R. Garcia and A. Zakhor. Geometric calibration for a multi-camera-projector system. In WACV, pages 467–474, 2013.
[11] R. I. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
[12] J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, pages 1106–1112, 1997.
[13] R. Horaud and F. Dornaika. Hand-eye calibration. International Journal of Robotics Research, 14(3):195–210, 1995.
[14] M. Kaess and F. Dellaert. Visual SLAM with a Multi-Camera Rig. Technical report, Georgia Institute of Technology, 2006.
[15] H. G. Maas. Image sequence based automatic multi-camera system calibration techniques. In International Archives of Photogrammetry and Remote Sensing, 32(B5):763–768, 1998.
[16] A. Malti. Hand-eye calibration with epipolar constraints: Application to endoscopy. Robotics and Autonomous Systems, 2012.
[17] E. Morais, A. Ferreira, S. A. Cunha, R. M. Barros, A. Rocha, and S. Goldenstein. A multiple camera methodology for automatic localization and tracking of futsal players. Pattern Recognition Letters, 2013.
[18] S. Shah and J. Aggarwal. Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation*. Pattern Recognition, 29(11):1775–1788, 1996.
[19] R. Tsai and R. Lenz. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. Robotics and Automation, IEEE Transactions on, 5(3):345–358, 1989.
[20] Z. Zhang. A flexible new technique for camera calibration. Technical report, Technical Report MSR-TR-98-71, Microsoft Research, 1998.
[21] Z. Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330–1334, 2000.
TABLE V
RESULT OF RELATIVE POSE CALIBRATION

III: our method. (see section VI-B) F: using feature correspondences.

| Algorithm | Relative Rotation (Degree) | Roll   | Pitch   | Yaw    |
|-----------|---------------------------|--------|---------|--------|
| III       |                           | 17.7158| -11.3660| -80.1913|
| F         |                           | 17.5459| -10.0024| -78.9854|

| Algorithm | Relative Translation (mm) | $T_x$   | $T_y$   | $T_z$   |
|-----------|---------------------------|---------|---------|---------|
| III       |                           | 295.4183| -232.4576| 34.5004|
| F         |                           | 294.0235| -229.8369| 28.9739|

Fig. 12
THE TRAJECTORY OF THE ACS RECOVERED FROM $Q_1$ AND $Q_2$

TABLE VI
RESULT OF RELATIVE POSE CALIBRATION USING NON-OVERLAPPING VIEW IMAGE SEQUENCES

III: our method. (see section VI-B) M: manual measurement

| Algorithm | Relative Rotation (Degree) | Roll   | Pitch   | Yaw    |
|-----------|---------------------------|--------|---------|--------|
| III       |                           | 1.3182 | 88.4530 | 0.7315 |
| M         |                           | 0 ± 5  | 90 ± 5  | 0 ± 5  |

| Algorithm | Relative Translation (mm) | $T_x$   | $T_y$   | $T_z$   |
|-----------|---------------------------|---------|---------|---------|
| III       |                           | 291.3321| -17.2837| -292.1382|
| M         |                           | 290±20  | 0±20    | 280±20  |

Fig. 13
THE ACS WITH TWO CANON POWERSHOT G9 USED IN THE REAL EXPERIMENT. (A) THE ACS USED IN THE REAL EXPERIMENT. (B) THE ACS AND TWO CHECKER PLANES. (C) IN THE FRONT OF THE ACS. (D) ON THE TOP OF THE ACS.
(a) Images Recorded by Camera A

(b) Images Recorded by Camera B

**Fig. 14**
Images Recorded by the ACS

**Fig. 15**
The trajectory of the ACS recovered from $Q_5$ and $Q_6$

**TABLE VII**
Result of Relative Pose Calibration with Unknown Scale Factors (0.8 in $Q_1$ and 3.2 in $Q_2$)

IV: our scale factor estimation method. (see section VI-C) F: using feature correspondences.

| Algorithm | Relative Rotation (Degree) | Roll  | Pitch | Yaw  |
|-----------|---------------------------|-------|-------|------|
| IV        |                           | 17.4883 | -10.5185 | -79.2551 |
| F         |                           | 17.5459 | -10.6024 | -78.9854 |

| Algorithm | Relative Translation (mm) | $T_x$ | $T_y$ | $T_z$ |
|-----------|--------------------------|-------|-------|-------|
| IV        |                          | 295.9218 | -220.6804 | 11.5566 |
| F         |                          | 294.0235 | -229.8369 | 28.9739 |
### TABLE VIII

| Algorithm | Relative Rotation (Degree) | Roll  | Pitch  | Yaw   |
|-----------|----------------------------|-------|--------|-------|
| Mean      | -4.4275                    | 38.9820 | -14.3572 |
| STD       | 0.4304                     | 0.2639 | 0.5774 |

| Algorithm | Relative Translation (mm) | $T_x$  | $T_y$  | $T_z$  |
|-----------|---------------------------|-------|--------|--------|
| Mean      | 489.2497                  | 56.0786 | -165.7425 |
| STD       | 6.2496                    | 3.1070 | 3.7616 |

| Algorithm | Relative Scale Factor | $\phi_{BA}$ |
|-----------|-----------------------|-------------|
| Mean      | 3.9531                |             |
| STD       | 0.0159                |             |