Stereo Calibration and Rectification for Omnidirectional Multi-camera Systems

Regular Paper

Yanchang Wang¹, Xiaojin Gong¹,*, Ying Lin¹ and Jilin Liu¹

1 Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang, China
* Corresponding author E-mail: gongxj@zju.edu.cn

Received 25 Apr 2012; Accepted 7 Jun 2012
DOI: 10.5772/50541

© 2012 Wang et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Stereo vision has been studied for decades as a fundamental problem in the field of computer vision. In recent years, computer vision and image processing with a large field of view, especially using omnidirectional vision and panoramic images, has been receiving increasing attention. An important problem for stereo vision is calibration. Although various kinds of calibration methods for omnidirectional cameras are proposed, most of them are limited to calibrate catadioptric cameras or fish-eye cameras and cannot be applied directly to multi-camera systems. In this work, we propose an easy calibration method with closed-form initialization and iterative optimization for omnidirectional multi-camera systems. The method only requires image pairs of the 2D target plane in a few different views. A method based on the spherical camera model is also proposed for rectifying omnidirectional stereo pairs. Using real data captured by Ladybug3, we carry out some experiments, including stereo calibration, rectification and 3D reconstruction. Statistical analyses and comparisons of the experimental results are also presented. As the experimental results show, the calibration results are precise and the effect of rectification is promising.

Keywords Stereo Calibration, Rectification, Omnidirectional Multi-camera System

1. Introduction

Stereo vision has been studied as a fundamental problem in computer vision for decades. By stereo matching, depth information can be obtained, which can be used in many applications such as autonomous navigation, obstacle detection, 3D reconstruction, virtual reality, object recognition and surveillance. For stereo matching, calibration and rectification are considered to be important pre-steps. By calibrating a stereo system, extrinsic parameters of the stereo system can be obtained to calculate the depth information. Calibration is also a must when rectifying stereo image pairs. A pair of images is thought to be rectified when its epipolar lines coincide [1]. One of the most important advantages of rectification is that the search for correspondences is done along the rows or the columns of the rectified images [2].

In recent years, computer vision and image processing with a large field of view (FOV), especially using omnidirectional vision and panoramic images, have been receiving increasing attention for navigation and surveillance applications [3]. Especially, the omnidirectional multi-camera system (OMS) has some more attractive advantages over other types of
omnidirectional cameras. Firstly, an OMS supplies higher resolution. Additionally, it also provides uniform resolution for any direction of view. Stereo systems composed of OMSs also have these advantages. Consequently, the research into the calibration and rectification methods for this kind of systems is meaningful and necessary.

Many algorithms for calibration and rectification have been proposed in omnidirectional applications. However, the existing approaches suffer the following defects: (1) Most of them are limited to calibrate and rectify catadioptric cameras with parabolic, hyperbolic, elliptical or spherical mirror and fisheye cameras. It is not convenient to apply them directly to OMSs (Ladybug3 [4], e.g.). (2) Some calibration methods need special facilities and scenes which are not simple to make and commonly available. (3) Some of the rectification methods are not scanline rectification methods, thus losing the important advantage of making the search for correspondences faster and simpler. (4) Part of the rectification methods produce heavily distorted images.

To solve these issues we propose a method for calibration and rectification in this paper which is proved to be well suited to OMSs. This calibration method is based on a spherical model without intrinsic parameters. We first deduce the projective properties of a group of equidistant parallel lines on the checkerboard. Based on these properties, the closed-form solutions of the extrinsic parameters are obtained. With the closed-form solutions as initial guesses, the calibration method produces a refined result by optimizing the object functions we propose. The rectification method we propose is a scanline method. It reprojects the stereo image pair onto a public unit sphere, on which the corresponding epipolar lines are projected onto the same longitude line. For a certain stereo system, the rectification process can be parameterized as a look-up table, which is easy to implement on a field-programmable gate array (FPGA) or a graphics processing unit (GPU).

There are two main contributions in this paper. Firstly, we propose a calibration method with iterative optimization and closed-form initialization which is well suited to OMSs. Secondly, we design a rectification algorithm [5] to project the corresponding epipolar lines onto the same longitude line on a unit sphere. This algorithm is a scanline algorithm and is convenient for hardware implementation with highly parallel processing. The algorithm also avoids the heavy distortion which is common in existing methods.

The remainder of this paper is organized as follows: In section 2 we introduce the related work. Section 3 discusses the proposed calibration method for an OMS. In section 4, we describe the process of stereo rectification in detail. Lastly we present our experimental results for real images and draw the conclusions.

2. Related Work

2.1 Calibration Methods for Omnidirectional Cameras

Researchers have proposed many calibration methods for single viewpoint omnidirectional cameras and most of these methods can be classified into the following three general categories [6]: the planar-grid-based methods, the line-based methods, and the non-prior-knowledge methods.

2.1.1 Planar-grid-based Calibration Methods

As methods using planar grids [7] are wildly adopted for their simplicity of use and accurate results in the calibration of perspective cameras, researchers naturally tend to develop similar approaches for omnidirectional cameras. D. Scaramuzza et al. [8] propose a method to calibrate single viewpoint omnidirectional cameras based on the assumption that the image projection function can be described by a Taylor series expansion whose coefficients are estimated by solving a two-step least squares linear minimization problem. However C. Mei and P. Rives point out that the user has to select each point of the calibration grid independently when calibrating for the use of the polynomial approximation of the projection function. They propose an improved method in [9] by introducing a slightly modified sphere model of Geyer [10] and Barreto [11]. This method is simple without the need to know the mirror parameters. Gasparini et al. recover the intrinsic parameters of the central catadioptric cameras from the image of the absolute conic (IAC) in [12]. Deng et al. [13] use the bounding ellipse and the field of view to initialize the intrinsic parameters and then obtain the extrinsic parameters by computing the homography using a direct linear transformation (DLT) algorithm with data normalization. Ikeda, S. et al. [14] propose a calibration method for an OMS. They compute the intrinsic and extrinsic parameters using a target plane and a laser total station. Generally speaking, the main advantages of this kind of method are the high precision, and that the target plane used is simple to make and commonly available.

2.1.2 Line-based Calibration Methods

Line-based calibration methods such as [15]-[19] usually need only one view to carry out the calibration process. These approaches estimate the intrinsic parameters of the catadioptric system by means of computing the image of the absolute conic. The main advantage of line-based calibration methods is that no special pattern is needed because lines are easily found in the natural scene. However, the accuracy of
this kind of method is usually lower than that of the planar-grid-based calibration methods.

2.1.3 Non-prior-knowledge Calibration Methods

Self-calibration methods, such as [20]-[22], use only point correspondences in multiple views, without needing any prior knowledge about the scene. However, they are less satisfactory for 3D reconstruction and motion estimation due to limited precision.

Compared with previous methods, our calibration method is also based on a sphere model, but the model is simplified without intrinsic parameters. Additionally, our method is a combination of the planar-grid-based method and the line-based method. The line-based algorithm is used in the initialization stage. As we mentioned above, this algorithm enables us to get relatively accurate initial calibration results conveniently from only one view. In the iterative optimization stage, we design the objective function using the planar-grid-based technique which produces higher precision.

2.2 Rectification Methods for Omnidirectional Cameras

Considering that former researchers have proposed a lot of effective methods on matching perspective stereo pairs, people try to rectify omnidirectional stereo pairs and then transform omnidirectional stereo problems into traditional stereo problems. Many methods are proposed to rectify different kinds of omnidirectional stereo image pairs. C. Geyer et al. propose a method to rectify stereo image pairs of parabolic catadioptric cameras [23]. However, this method is limited to parabolic catadioptric stereo systems. Also, the method produces heavily distorted images. S. Abraham et al. propose an algorithm for fish-eye-stereo calibration and epipolar rectification [3]. Although it avoids heavy distortion and over expansion near the epipoles, the method is not a scanline rectification method, thus losing the important advantage of making computing stereo correspondences simpler. Researchers also exploit the rectification under different geometry frameworks. For example, Gonzalez-Barbosa et al. [24] rectify omnidirectional images on the rectangular grid, while Takiguchi et al. [25] reproject omnidirectional images onto cylinders. F. Kangni et al. [26] rectify the image pairs using the cubic projection model.

Although all the approaches mentioned above exhibit good behaviour in rectification, they still suffer at least one of the following defects: (1) some of the rectification methods are not scanline rectification methods, thus losing the important advantage of making the search for correspondences faster and simpler; (2) some of them produce heavily distorted images; (3) some are not easy for hardware implementation.

3. Stereo Calibration of OMs

3.1 The Ladybug3 Omnidirectional Camera System

The high resolution Ladybug3 spherical digital video camera system () has six 2 MP cameras that enable the system to collect video from more than 80% of the full 360° sphere, and an IEEE-1394b (FireWire) interface with locking screw connection that allows JPEG-compressed 12MP resolution images to be streamed to disk at 15fps. The Ladybug3 camera system can output not only six normal images, but also an omnidirectional image either in DOME mode or Mercator mode. The Mercator mode panoramic images provided by the Ladybug3 are created using a spherical coordinate system. The projection model is as follows. The image axes u and v map to the two spherical angles θ and φ respectively in the following formulations:

\[
\begin{aligned}
    u &= \frac{\pi - \theta}{2\pi} \cdot n\text{Cols} \\
    v &= \frac{\phi}{\pi} \cdot n\text{Rows},
\end{aligned}
\]  

where \(n\text{Cols} \times n\text{Rows}\) is the size of the panoramic images, \(\theta\) and \(\phi\) defined in are computed as follows:

\[
\begin{aligned}
    \theta &= \text{atan2}(y,x) \\
    \phi &= \arccos \frac{z}{\sqrt{x^2 + y^2 + z^2}},
\end{aligned}
\]

where \((x,y,z)\) is a point in the camera head coordinate frame. Substitute Eq. 2 into Eq. 1, we get

\[
\begin{aligned}
    u &= \frac{\pi - \text{atan2}(y,x)}{2\pi} \cdot n\text{Cols} \\
    v &= \frac{\arccos(z/d)}{\pi} \cdot n\text{Rows},
\end{aligned}
\]

where \(d = \sqrt{x^2 + y^2 + z^2}\).

**Figure 1.** Ladybug3 Omnidirectional System.
3.2 Independent Extrinsic Calibration

In this part, we will calibrate the rotation and the translation between the target plane and the camera in each view. The final results are solved with the Levenberg-Marquardt algorithm. Consequently, we will discuss the initial guess of the rotation and the translation first. And then, a complete procedure of estimating the best result will be given.

3.2.1 Initialization of Extrinsic Parameters in Each View

In our algorithm of initialization, we will first estimate the rotation matrix. Then the translation can be obtained based on the estimated rotation matrix.

Let \( \mathbf{P}_w \) be the coordinate of an arbitrary point attached to the target plane in the world reference frame and \( \mathbf{R}_{wc} \) be the \( 3 \times 3 \) orthonormal rotation matrix that rotates the world reference frame into the camera head coordinate frame and \( \mathbf{T}_{wc} \) be the the translation. Then the transformation equation that transforms a point from the world reference frame to the camera head coordinate frame can be written as:

\[
\mathbf{P}_c = \mathbf{R}_{wc}\mathbf{P}_w + \mathbf{T}_{wc}.
\]  

(4)

Without loss of generality, we assume the model plane is on the \( x - y \) plane of the world coordinate system, as shown in , that is \( z_w = 0 \). Let \( dx \) and \( dy \) denote the size of each square in the grid in \( X_w \) and \( Y_w \).

There are two sets of parallel lines on the checkerboard. Each line is perpendicular to the lines in the other set. When projected to the viewing sphere in , each line forms a great circle. As shown in , the great circles corresponding to the parallel lines from the same set intersect at two antipodal junctions (\( \mathbf{E}_{x^+} \), \( \mathbf{E}_{x^-} \) and \( \mathbf{E}_{y^+} \), \( \mathbf{E}_{y^-} \)), just like the meridian lines intersect at the North Pole and the South Pole. It is easy to prove that vector \( \mathbf{OE}_{x^+} \) is parallel with axis \( X_w \) and vector \( \mathbf{OE}_{y^+} \) is parallel with axis \( Y_w \). Therefore, if we denote the \( i \)th column of the rotation matrix \( \mathbf{R}_{wc} \) by \( \mathbf{r}_i \), we have

\[
\begin{align*}
\mathbf{r}_1 &= \mathbf{E}_{x^+} \\
\mathbf{r}_2 &= \mathbf{E}_{y^+} \\
\mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2
\end{align*}
\]  

(5)

Once we obtain the initial guess of the rotation matrix \( \mathbf{R}_{wc} \), we are able to solve the estimated translation vector \( \mathbf{T}_{wc} \). Let \( \mathbf{P}_{c,mn} = (u_{mn}, v_{mn}) \) be the pixel of a grid corner on the checkerboard, where the subscripts \( m \) and \( n \) denote the indexes of the row and the column. According to Eq. 1, we have \( \theta_{mn} = \pi - 2\pi \cdot u_{mn}/\text{Cols} \) and \( \phi_{mn} = \pi \cdot v_{mn}/\text{Rows} \). As shown in , we define \( \mathbf{OP} \) \((\theta, \phi, ||\mathbf{OP}||)\) as the unit directional vector of \( \mathbf{OP} \). For any grid corner \( \mathbf{P}_{c,mn} \) on the checkerboard, we can write the unit directional vector of \( \mathbf{OP}_{c,mn} \) as

\[
M_{mn} = \mathbf{OP}_{c,mn}(\theta_{mn}, \phi_{mn}, 1) =
\begin{bmatrix}
\sin\phi_{mn}\cos\theta_{mn} \\
\sin\phi_{mn}\sin\theta_{mn} \\
\cos\phi_{mn}
\end{bmatrix}.
\]  

(6)
Let $d_{mn}$ be the distance from $P_{c,mn}$ to the center of the viewing sphere, that is $d_{mn} = \|OP_{c,mn}\|$. We have

$$P_{c,mn} = d_{mn}M_{mn}. \quad (7)$$

As the normal of the checkerboard is $r_3$, assuming that the distance from the center of the viewing sphere to the checkerboard plane is $d_c$, any point $P_{c,mn}$ on the checkerboard plane satisfies the following equation

$$r_3 \cdot P_{c,mn} + d_c = 0. \quad (8)$$

Substitute Eq. 7 into Eq. 8, we get

$$r_3 \cdot (d_{mn}M_{mn}) + d_c = 0. \quad (9)$$

We can rewrite it as

$$d_{mn} = -\frac{d_c}{r_3 \cdot r_{mn}}. \quad (10)$$

Substitute Eq. 10 into Eq. 7, we get

$$P_{c,mn} = -\frac{d_c}{r_3 \cdot r_{mn}}M_{mn}. \quad (11)$$

Let $P_{w,mn}$ be the corresponding coordinate of $P_{c,mn}$ in the world coordinate system. According to Eq. 4, the relationship between them is given by:

$$P_{w,mn} = R_{wc}^T P_{c,mn} - R_{wc}^T T_{wc}. \quad (12)$$

Since the model plane is on the $x-y$ plane of the world coordinate system and the point $P_{w,11}$ is the origin, as shown in , we have $P_{w,11} = 0$. Substitute it into Eq. 4, we have $P_{c,11} = R_{wc}P_{w,11} + T_{wc}. \quad (13)$

Substitute Eq. 13 and Eq. 11 into Eq. 12, we have

$$P_{w,mn} = R_{wc}^T P_{c,mn} - R_{wc}^T T_{wc} = \begin{bmatrix} r_1 \cdot M_{11} - r_1 \cdot M_{mn} \\ r_2 \cdot M_{11} - r_2 \cdot M_{mn} \\ r_3 \cdot M_{11} - r_3 \cdot M_{mn} \end{bmatrix} d_c \quad (14)$$

According to , we can rewrite $P_{w,mn}$ as

$$P_{w,mn} = \begin{bmatrix} (m-1)dX \\ (n-1)dY \\ 0 \end{bmatrix}. \quad (15)$$

From Eq. 14 and Eq. 15, we can construct two equations with a single point $P_{w,mn}$:

$$\begin{cases} (r_1 \cdot M_{11} - r_1 \cdot M_{mn}) d_c = (m-1)dX \\ (r_2 \cdot M_{11} - r_2 \cdot M_{mn}) d_c = (n-1)dY \end{cases} \quad (16)$$

For a checkerboard with $M \times N$ grid corners, we can construct $(M-1)N + M(N-1)$ equations. All these equations form an overdetermined system with the form $A\hat{d}_c = b$. The least squares formula can be written as

$$\hat{d}_c = (A^T A)^{-1}A^T b, \quad (17)$$

where

$$A^T A = \sum_{m=2}^{M} \sum_{n=1}^{N} \left( r_1 \cdot M_{11} \cdot M_{mn} - r_1 \cdot M_{mn} \right)^2 \quad (18)$$

and

$$A^T b = \sum_{m=2}^{M} \sum_{n=1}^{N} \left( r_1 \cdot M_{11} \cdot M_{mn} - r_1 \cdot M_{mn} \right) (m-1)dX \quad (19)$$

According to Eq. 13 and Eq. 11, we have

$$T_{wc} = \frac{d_c}{r_3 \cdot r_{m11}} M_{11}. \quad (20)$$

3.2.2 Optimization of Extrinsic Parameters

Next, we will find the final results of the extrinsic parameters by applying the iterative optimization.

By extracting the grid corners in the images, we are given $M \times N$ points on the target plane. Assume that the image points are corrupted by independent and identically distributed noise. The maximum likelihood estimate of $R_{wc}$ and $T_{wc}$ can be obtained by minimizing the following function:

$$R_{wc}, T_{wc} = \arg \min \sum_{m=2}^{M} \sum_{n=1}^{N} \|P_{c,mn} - \tilde{P}_{c,mn}(R_{wc}, T_{wc}, P_{w,mn})\|^2, \quad (21)$$

where $\tilde{P}_{c,mn}(R_{wc}, T_{wc}, P_{w,mn})$ is the projection of point $P_{w,mn}$ according to Eq. 3 and Eq. 4. For a known target plane, $P_{w,mn}$ can be easily calculated from Eq. 15 and the corresponding projection $P_{c,mn}$ can be obtained by extracting the grid corners in the image. This is a nonlinear minimization problem, which can be solved with the Levenberg-Marquardt Algorithm [27][28]. After we get $R_{wc}$ and $T_{wc}$, the normal of the target plane in the camera frame can be calculated. Let’s rewrite Eq. 4 as follows:

$$\begin{cases} \left( r_1 \cdot M_{11} - r_1 \cdot M_{mn} \right) d_c = (m-1)dX \\ \left( r_2 \cdot M_{11} - r_2 \cdot M_{mn} \right) d_c = (n-1)dY \end{cases} \quad (16)$$
\[ \mathbf{R}_{\text{wc}} \mathbf{p}_c - \mathbf{R}_{\text{wc}} \mathbf{T}_{\text{wc}} = \mathbf{p}_w \]  
(22)

Noticing that \( \mathbf{p}_w = [x_w \ y_w \ 0]^T \), we have
\[ \mathbf{r}_3 \cdot \mathbf{p}_c - \mathbf{r}_3 \cdot \mathbf{T}_{\text{wc}} = 0. \]  
(23)

Therefore, the normal of the target plane in the camera frame is given by:
\[ \mathbf{N} = (\mathbf{r}_3 \cdot \mathbf{T}_{\text{wc}}) \mathbf{r}_3. \]  
(24)

### 3.3 Stereo System Calibration

We treat the stereo system as a strictly rigid system. Consequently, the calibration of this stereo system is to estimate the best transform between Camera \( h \) and Camera \( l \). We propose a two-step calibration process. In step I, we decouple the problem of finding the best transform into estimating a rotation part and a translation part independently. In step II, we use the results in step I as the initial guess to optimize the cost function to refine the results. We have two choices of the cost function: the first one is chosen to be the distance from the 3D points of the grid corners in one camera to the corresponding target plane observed from the other camera; the second one is the Euclidean distance between the 3D correspondences observed in the two cameras.

For each posture of the target plane, we can calculate its rotation \( \mathbf{R}_{\text{wc}} \), transformation \( \mathbf{T}_{\text{wc}} \) and the normal of the target plane \( \mathbf{N} \) in the camera head reference frame of Camera \( h \) and Camera \( l \) from the previous section. Let’s denote them as \( \mathbf{R}_{h,i} \), \( \mathbf{T}_{h,i} \), \( \mathbf{N}_{h,i} \) and \( \mathbf{R}_{l,i} \), \( \mathbf{T}_{l,i} \), \( \mathbf{N}_{l,i} \), respectively, where the subscripts \( h \) and \( l \) denote the camera head reference frame \( h \) and \( l \); and \( i \) is the index of the observation. Let
\[
\mathbf{N}_h = \begin{bmatrix} \mathbf{N}_{h,1} & \mathbf{N}_{h,2} & \cdots & \mathbf{N}_{h,m} \end{bmatrix},
\]
\[
\mathbf{D}_h = \begin{bmatrix} \mathbf{N}_{h,1} & \mathbf{N}_{h,2} & \cdots & \mathbf{N}_{h,m} \end{bmatrix}^T,
\]  
(25)

and
\[
\mathbf{N}_l = \begin{bmatrix} \mathbf{N}_{l,1} & \mathbf{N}_{l,2} & \cdots & \mathbf{N}_{l,m} \end{bmatrix},
\]
\[
\mathbf{D}_l = \begin{bmatrix} \mathbf{N}_{l,1} & \mathbf{N}_{l,2} & \cdots & \mathbf{N}_{l,m} \end{bmatrix}^T,
\]  
(26)

where \( m \) is the number of stereo image pairs.

Then the estimate of the translation has a closed form solution as follows:
\[ \mathbf{T}_{hl} = (\mathbf{N}_l \mathbf{N}_l^T)^{-1} \mathbf{N}_l (\mathbf{D}_1 - \mathbf{D}_h)^T, \]  
(27)

and
\[ \mathbf{R}_{hl} = \mathbf{VU}^T, \]  
(28)

where \( \mathbf{N}_l \mathbf{N}_l^T = \mathbf{USV}^T \) is the associated singular value decomposition and the subscript \( hl \) means the translation is from the reference frame \( h \) to the reference frame \( l \).

In Step II, what we do is minimizing a cost function which stands for the error of \( \mathbf{R}_{hl} \) and \( \mathbf{T}_{hl} \) to reduce the influence of noise. Here we introduce two kinds of cost functions used for optimization. The minimization of both the objective functions can be done through an iterative optimization procedure with initial estimates obtained from step I.

Let \( \mathbf{P}_{h,i}^{l} \) be the \( j \)th 3D grid point from the \( i \)th image captured by Camera \( h \) and \( \mathbf{P}_{l,i}^{l} \) be the corresponding points from the image of Camera \( l \). For \( m \) different views of the target plane and \( n \) 3D grid corner points per view, the refined extrinsic parameters \( \mathbf{R}_{hl} \) and \( \mathbf{T}_{hl} \) can be obtained by minimizing the following reprojection error:
\[
E = \sum_{i} \sum_{j} \frac{\mathbf{N}_{l,i}}{\left\| \mathbf{N}_{l,i} \right\|} \cdot \left( (\mathbf{R}_{h,i} \mathbf{P}_{h,i}^{l} + \mathbf{T}_{h,i}) - \mathbf{P}_{l,i}^{l} \right)^2. \]  
(29)

This means the distance from the 3D points of the grid corners in one camera to the corresponding target plane observed from the other camera should be zero if the extrinsic parameters \( \mathbf{R}_{hl} \) and \( \mathbf{T}_{hl} \) are ideally estimated.

Another objective function is given more naturally in that it requires the Euclidean distance between the 3D points observed in the two cameras should be a minimum. The objective function is given as follows:
\[
E = \sum_{i} \sum_{j} \left\| (\mathbf{R}_{h,i} \mathbf{P}_{h,i}^{l} + \mathbf{T}_{h,i}) - \mathbf{P}_{l,i}^{l} \right\|^2. \]  
(30)

These two nonlinear optimization problems given in Eq. 29 and Eq. 30 can be solved by using the Levenberg-Marquardt algorithm.

When solving the optimization problem using the first object function Eq. 29, we should notice that at least three views of the target plane which are nonparallel to each other are required to fully constrain the optimization problem. Since the solution of the system has 6 degrees of freedom (DOF) and each view can constrain 2 DOF, we need at least three views of the target plane. If the number of views is smaller than three, the solution obtained will not converge to the actual value.
4. Stereo Rectification Method

4.1 Geometric Framework

Firstly, let’s make some notations clear. Let \( O_h \) be the origin of the camera head reference frame \( h \) and \( O_l \) be the the origin of the camera head reference frame \( l \). From the calibration data, we are able to derive the extrinsic parameters \( R_{hl} \) and \( T_{hl} \) of the omnidirectional stereo camera system, in which \( R_{hl} \) is the \( 3 \times 3 \) rotation matrix from \( h \) to \( l \) and \( T_{hl} \) is the translation.

Imagining that if we set up a rectification reference frame \( r \) whose \( z \) is along the baseline \( T_{hl} \), as shown in Figure 5, every epipolar plane is perpendicular to the \( x - y \) plane. It means that all the points belonging to a certain epipolar plane are projected on the same line in the \( x - y \) plane. Consequently, all of the points on the same epipolar line should have the same ratio of their \( x \) and \( y \) coordinates (aligned in frame \( r \)). In other words, an epipolar plane intersects two spheres at two great circles with constant zenithal angles if we consider the epipoles as the two poles of the spheres.

In order to transform the coordinates from the reference frames \( h \) and \( l \) to the reference frame \( r \), we have to decide \( R_{hr} \) and \( R_{lr} \), which are the rotation matrices from the reference frames \( h \) and \( l \) to the reference frame \( r \) respectively. Next we will show that \( R_{hr} \) is able to be derived as long as \( R_{lr} \) is known. As shown in Figure 5, let \( P \) be an arbitrary point in 3D space. The coordinates of point \( P \) in the reference frame of camera system \( h, l \) and \( r \) are \( P_h, P_l \) and \( P_r \) respectively. We have

\[
\begin{align*}
P_r &= R_{lr} P_l + T_{lr} \\
&= R_{hr} P_h + T_{hr},
\end{align*}
\]

where \( T_{lr} = 0 \) for the origins of the reference frame \( l \) and \( r \) are coincident. As a rigid transformation, we have

\[
P_l = R_{hl} P_h + T_{hl}.
\]

By substituting Eq. 32 and \( T_{lr} = 0 \) into Eq. 31, we get

\[
\begin{align*}
P_r &= R_{lr} R_{hl} P_h + R_{lr} T_{hl} \\
&= R_{hr} P_h + T_{hr},
\end{align*}
\]

From Eq. 33, we can derive

\[
\begin{align*}
R_{hr} &= R_{lr} R_{hl} \\
T_{hr} &= R_{lr} T_{hl}.
\end{align*}
\]

Obviously, once \( R_{lr} \) is determined, transformation from the reference frames \( h \) and \( l \) to the reference frame \( r \) is able to be computed.

Next, let’s estimate \( R_{lr} \). As shown in the previous section, \( R_{hl} \) and \( T_{hl} \) are known after calibration. Considering point \( O_h \), its coordinate vector in reference frame \( l \) is vector \( T_{hl} \). At the same time, point \( O_h \) is on the \( z \) axis of the reference frame \( r \) and its coordinates in the reference frame \( r \) are \([0,0,||T_{hl}||]^T \). \( R_{lr} \) is the \( 3 \times 3 \) rotation matrix and we write it as follows:

\[
R_{lr} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}.
\]

Substitute the coordinates of point \( O_h \) in the reference frames \( h \) and \( r \) into Eq. 33, we have

\[
[0,0,||T_{hl}||]^T = R_{lr} T_{hl}
\]

and the following constraint because \( R_{lr} \) is orthonormal as a rotation matrix:

\[
R_{lr} R_{lr}^T = I,
\]

where \( I \) is an identity matrix. Eq. 36 and Eq. 37 form a group of indeterminate equations because the reference frame \( r \) can rotate around the axis \( z \). For simplicity, we let \( r_{11} = 0 \) and show the corresponding analytic solutions as follows:

Figure 5. All the points belonging to a certain epipolar plane are projected on the same line in the \( x - y \) plane. For clarity, the \( x - y \) plane is moved from \( O_r \) to \( O_r' \).
\[ r_{12} = -\frac{r_{33}}{r_{32}} \sqrt{\frac{r_{22}^2}{r_{32}^2 + r_{33}^2}}, \]
\[ r_{13} = \frac{r_{32}^2}{r_{32}^2 + r_{33}^2}, \]
\[ r_{21} = -\frac{r_{22}(r_{32}^2 + r_{33}^2)}{r_{31}r_{32}}, \]
\[ r_{22} = \frac{1}{\sqrt{\frac{r_{32}^2}{r_{32}^2 + r_{33}^2} + \frac{r_{33}^2}{r_{32}^2 + r_{33}^2} + 1}}, \]
\[ r_{23} = \frac{r_{33}}{r_{32}}, \]
\[ [r_{31} \ r_{32} \ r_{33}] = \frac{T_{l}^T}{\|T_{l}\|} \]

\[ (38) \]

### 4.2 Rectification with Interpolation

Without loss of generality, we do not limit the projection model of the camera as long as the correspondences in the same row or column of the rectified images have constant zenithal angles if we take frame \( r \) as the reference for the spherical coordinate system. Here we take Mercator projection as the model and the relation between every pixel in the rectified images and the corresponding points on the sphere is given by Eq. 3.

Supposing that we want to derive the intensity of a pixel with the coordinate \([i, j]\) in the rectified image, we first calculate its coordinates on the sphere \( X_i \) in the frame \( r \) by transforming its spherical polar coordinate \((d, \theta, \varphi)\) into a Cartesian coordinate \( X_{ir} \), where \( d \) is the radius of the sphere, \( \theta \) and \( \varphi \) can be obtained by Eq. 1. Thus its Cartesian coordinate \( X_i \) in frame \( l \) is \( X_i = R_{ir}^T X_{ir} \), where \( R_{ir}^T = R_{ir} \) because the rotation matrix \( R_{ir} \) is orthonormal. Then we can compute its corresponding pixel coordinates \([m, n]\) in the image of Camera \( l \) by Eq. 3. In most cases \( m \) and \( n \) are not integers. This means that the corresponding pixels fall into the rectangular region delimited by the four original pixels in the original images. In order to derive a more accurate result of the rectified image, some interpolation techniques need to be introduced. For the balance of complexity and performance, we choose bilinear interpolation as our preferred interpolation method. Let \( I_l(i,j) \) be the intensity of pixel \([i,j]\) in the rectified images and \( I_o([m],[n]) \) be the intensity of pixel \([m],[n]\) in the original images, we have

\[ I_l(i,j) = I_o([m],[n]) (1 - p)(1 - q) + I_o([m],[n])p(1 - q) + I_o([m],[n] + 1)(1 - p)q + I_o([m] + 1,[n] + 1)pq, \]

where \( p = m - [m] \) and \( q = n - [n] \). \([\bullet]\) means rounding the element to the nearest integer towards minus infinity.

### 5. Experimental Results

The experiments are conducted on the platform of Ladybug3. Camera \( h \) and Camera \( l \) are installed with a baseline of about 37 cm in the vertical direction. The resolution of the images captured by Ladybug3 can reach up to 5,400×2,700, which enables us to obtain a more precise calibration result. We carry out the experiments in two different scenes and show one of them for rectification and the other for 3D reconstruction.

#### 5.1 Calibration

The method we proposed requires image pairs of the 2D target plane in a few different views. In this experiment, a target plane with a printed chessboard pattern is used. The target plane has \( 11 \times 11 \) squares and the side of each square is 30 mm. We get nine pairs of stereo images in total. In this experiment, we have done some statistical analysis on the calibration results. The rotation part of the calibration result is expressed in Euler angles \((\alpha, \beta, \gamma)\), and the translation part is expressed as a 3D vector \((x_t, y_t, z_t)\). In Table 1, we have shown the variation of the calibration results obtained from different numbers of images. The results shown in Table 1 are obtained by using Eq. 29 and Eq. 30. As the method using Eq. 29 requires at least three views of the target plane to give the right results, we set the ‘Num of images’ to increase from 3. We can see that the \( \sigma \) of the results given by Eq. 29 is generally larger than that produced by Eq. 30.

In order to investigate the stability of the proposed methods, we have applied the two methods using Eq. 29 and Eq. 30 to all combinations of eight images from the available nine images. We use Exc. \( i \) to denote the combination with the \( i \)th image excluded. For example, Exc. 9 means the results come from the images of 1-8. All the results are shown in Table 2 and Table 3. We find that the deviations from both methods are quite small and it implies that both the proposed algorithms can be considered stable. Since the deviation of the results given by Eq. 29 is still slightly larger, this may indicate that the method using Eq. 30 is probably more robust.

We also compare the results from our method with other previous methods in Table 4. The existing algorithms or toolboxes are mostly designed specifically for dioptric or catadioptric cameras. Their contributions are mainly concerned with the estimation of the parameters of the projection model. The Ladybug3 system is not only an OMS without intrinsic parameters, but also with a different projection model. In order to apply the previous methods on the Ladybug3 system, we partly modify the methods, including removing the calculation of the intrinsic parameters and replacing the projection model. We compare two kinds of the existing methods with ours
Table 1. Results using real data of 3 to 9 images using Eq. 29 and Eq. 30

<table>
<thead>
<tr>
<th>Num of images</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>-3.92</td>
<td>-2.12</td>
<td>-0.09</td>
<td>-0.72</td>
<td>-1.25</td>
<td>-0.45</td>
<td>8.30</td>
</tr>
<tr>
<td>Final</td>
<td>-3.92</td>
<td>-2.19</td>
<td>-1.07</td>
<td>-1.10</td>
<td>-9.92</td>
<td>-7.52</td>
<td>5.0</td>
</tr>
<tr>
<td>Initial σ</td>
<td>4.80</td>
<td>3.10</td>
<td>2.20</td>
<td>1.10</td>
<td>10.70</td>
<td>8.30</td>
<td>6.30</td>
</tr>
<tr>
<td>Final σ</td>
<td>5.50</td>
<td>3.60</td>
<td>3.60</td>
<td>4.80</td>
<td>3.60</td>
<td>4.40</td>
<td>3.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exc. 1</th>
<th>Exc. 2</th>
<th>Exc. 3</th>
<th>Exc. 4</th>
<th>Exc. 5</th>
<th>Exc. 6</th>
<th>Exc. 7</th>
<th>Exc. 8</th>
<th>Exc. 9</th>
<th>Mean</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>α(10⁻³)</td>
<td>-1.54</td>
<td>-2.93</td>
<td>-0.19</td>
<td>-1.25</td>
<td>10.70</td>
<td>8.30</td>
<td>6.30</td>
<td>354.83</td>
<td>355.88</td>
<td>354.10</td>
</tr>
<tr>
<td>β(10⁻³)</td>
<td>2.93</td>
<td>2.89</td>
<td>0.19</td>
<td>1.10</td>
<td>8.30</td>
<td>6.30</td>
<td>354.83</td>
<td>355.88</td>
<td>354.10</td>
<td>354.10</td>
</tr>
<tr>
<td>γ(10⁻³)</td>
<td>-0.28</td>
<td>-1.62</td>
<td>0.29</td>
<td>-1.25</td>
<td>-9.92</td>
<td>-7.52</td>
<td>354.83</td>
<td>355.88</td>
<td>354.10</td>
<td>354.10</td>
</tr>
<tr>
<td>tₓ(mm)</td>
<td>1.80</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
<td>2.81</td>
</tr>
<tr>
<td>t-svg(mm)</td>
<td>354.83</td>
<td>355.88</td>
<td>354.10</td>
<td>354.10</td>
<td>353.61</td>
<td>355.80</td>
<td>350.97</td>
<td>352.33</td>
<td>354.17</td>
<td>354.20</td>
</tr>
</tbody>
</table>

Table 2. Variation of the calibration results among all continuous 8 images using Eq. 29

<table>
<thead>
<tr>
<th>Method</th>
<th>α(10⁻³)</th>
<th>β(10⁻³)</th>
<th>γ(10⁻³)</th>
<th>tₓ(mm)</th>
<th>tᵧ(mm)</th>
<th>t-svg(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method (Eq. 29)</td>
<td>-7.52</td>
<td>2.46</td>
<td>-0.42</td>
<td>-1.25</td>
<td>-96.40</td>
<td>353.82</td>
</tr>
<tr>
<td>Our Method (Eq. 30)</td>
<td>-7.52</td>
<td>2.46</td>
<td>-0.42</td>
<td>-1.25</td>
<td>-96.40</td>
<td>353.82</td>
</tr>
<tr>
<td>PGB (Eq. 29)</td>
<td>1.68</td>
<td>83.85</td>
<td>49.61</td>
<td>31.64</td>
<td>-93.03</td>
<td>364.07</td>
</tr>
<tr>
<td>PGB (Eq. 30)</td>
<td>-2.79</td>
<td>75.60</td>
<td>43.54</td>
<td>28.48</td>
<td>-94.54</td>
<td>358.01</td>
</tr>
<tr>
<td>LB (Eq. 29)</td>
<td>-1.88</td>
<td>3.40</td>
<td>2.85</td>
<td>0.33</td>
<td>-101.90</td>
<td>355.06</td>
</tr>
<tr>
<td>LB (Eq. 30)</td>
<td>3.26</td>
<td>2.09</td>
<td>-0.18</td>
<td>0.84</td>
<td>-101.10</td>
<td>356.66</td>
</tr>
</tbody>
</table>

Table 3. Variation of the calibration results among all continuous 8 images using Eq. 30

<table>
<thead>
<tr>
<th>Method</th>
<th>α(10⁻³)</th>
<th>β(10⁻³)</th>
<th>γ(10⁻³)</th>
<th>tₓ(mm)</th>
<th>tᵧ(mm)</th>
<th>t-svg(mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method (Eq. 29)</td>
<td>-7.52</td>
<td>2.46</td>
<td>-0.42</td>
<td>-1.25</td>
<td>-96.40</td>
<td>353.82</td>
</tr>
<tr>
<td>Our Method (Eq. 30)</td>
<td>-7.52</td>
<td>2.46</td>
<td>-0.42</td>
<td>-1.25</td>
<td>-96.40</td>
<td>353.82</td>
</tr>
<tr>
<td>PGB (Eq. 29)</td>
<td>1.68</td>
<td>83.85</td>
<td>49.61</td>
<td>31.64</td>
<td>-93.03</td>
<td>364.07</td>
</tr>
<tr>
<td>PGB (Eq. 30)</td>
<td>-2.79</td>
<td>75.60</td>
<td>43.54</td>
<td>28.48</td>
<td>-94.54</td>
<td>358.01</td>
</tr>
<tr>
<td>LB (Eq. 29)</td>
<td>-1.88</td>
<td>3.40</td>
<td>2.85</td>
<td>0.33</td>
<td>-101.90</td>
<td>355.06</td>
</tr>
<tr>
<td>LB (Eq. 30)</td>
<td>3.26</td>
<td>2.09</td>
<td>-0.18</td>
<td>0.84</td>
<td>-101.10</td>
<td>356.66</td>
</tr>
</tbody>
</table>

Table 4. Comparison of our calibration method with other previous ones
In Table 4. One is the planar-grid-based (PGB) method similar to [9], in Table 4. The other is the line-based (LB) method similar to [15]. The PGB and LB methods in Table 4 are followed by Eq. 29 or Eq. 30, which means Eq. 29 or Eq. 30 is chosen as the object function for the optimization of the stereo calibration.

5.2 Stereo Rectification

Figure 6 is the original image pair captured by Ladybug3 in this experiment and Figure 7 shows the rectified images from our rectification method. We line up the correspondences which are found and matched using Scale-invariant feature transform (SIFT). According to Figure 7, we can see that the correspondences fall on the same columns in the rectified pair. In order to make the results clear, we make some comparisons between the picture blocks from an unrectified stereo pair and a rectified stereo pair.

Figure 8(a) presents the details from the unrectified image pair. It is clearly shown that the stereo correspondences are not on the same line. Consequently, the performance of stereo matching will be poor.

Figure 8(b) presents the details from the rectified image pair. We can observe that the stereo correspondences fall on the same line after rectification. Comparing the details from Figure 8(a) and Figure 8(b), we can conclude that our method of calibration and rectification is effective.

We also record the time required to perform the rectification on a PC platform. The average time to rectify a 2700×1350 grey image is 2.66 seconds, while the average time for a coloured image of the same size is 9.29 seconds. All experiments were carried out on a server with an Intel Xeon X5560 CPU (2.80GHz), 12G RAM, running Matlab 2009 64bit. For an FPGA with an 80MHz system clock, it theoretically takes 228ms to rectify a 2700×1350 grey image pair by taking advantage of the parallel computing and the pipeline framework. For a coloured image of the same size, the time increases to 0.68s.

Figure 6. Original image pair

Figure 7. Rectified image pair

Figure 8. Comparison between the picture blocks from an unrectified stereo pair and a rectified stereo pair. Picture blocks in the first row are from Camera h and those in the second row are from Camera l.
5.3 3D reconstruction

In this experiment, we have also obtained the 3D reconstruction results of the example scene by calculating the intersection of two rays using equation [29]. The virtual bird’s-eye view (by IPM) is shown in (a). The reconstruction results can be seen in (b). As the experiments are carried out in the corner of the room, we should observe two walls perpendicular to each other in the bird’s-eye view of the reconstruction result. (b) strongly agrees with this inference. In the point cloud, the two perpendicular walls, the desks and the chair can be clearly seen. We have obtained the 360° 3D reconstruction results of the example scene and the results are promising.

6. Conclusion

In this paper, we have developed an easy stereo calibration and rectification method for omnidirectional multi-camera systems. In the calibration procedure, iterative optimization is adopted to reduce the influence of noise on the calibration results. We give the closed-form solution based on the projective properties of equidistant parallel lines as the initial guess for iterative optimization. We have tested the algorithm with real data and carried out some analysis on the calibration results. According to the analysis, we believe that the method we propose is robust and reliable. Furthermore, we have applied the scanline rectification method we propose to the unrectified images. By comparing the details before and after rectification, we can conclude that the calibration results are precise enough to support rectification and that the rectification method is well suited to the OMS. Generally, we can conclude that the calibration and the rectification methods for the OMS are effective and promising.

7. Acknowledgements

This research work was supported in part by the National Natural Science Foundation of China via grant 61001171, 60534070 and 90820306, and by the Chinese Universities Scientific Fund.

8. References


