Comparison of Two Panoramic Sensor Models for Precise 3D Measurements

Shunping Ji, Yun Shi, Zhongchao Shi, Anmin Bao, Junli Li, Xiuxiao Yuan, Yulin Duan, and Ryosuke Shibasaki

Abstract

In this paper, the system errors produced by the most widely used ideal panoramic camera model for a close-range multi-camera rig are indicated and analytically modeled according to a rigorous panoramic camera model, and a comprehensive comparison between the two models is given. First, the 3D localization errors of the ideal model are analyzed that shows the correlations with the object-image distance and the viewing angles. Second, the epipolar errors are analyzed and are observed to exhibit changes with the rotation angles and z-coordinates of the image points. Finally, tests are carried out in space resection, epipolar constraints, and bundle adjustment with different sensor models. The outdoor tests with small object-image distance (several meters) show the difference between the two models is notably slight. In contrast, the indoor tests with larger object-image distance (more than 15 m) show the rigorous model produces 2 cm better measurement accuracy than the ideal.

Introduction

In many research studies and applications in recent years (Li *et al.*, 2004; Anguelov *et al.*, 2010), the close-range panoramic camera has been employed in place of a traditional plane camera because it features full panoramic information in a single image and a simple structure: one projection center and one projection sphere or cylinder. However, compared with a plane camera, larger geometric distortions exist in a panoramic camera or in a fish-eye camera even with a much smaller field of view, which may result in poor imaging quality. From a manufacturing perspective, there are three main methods used to overcome the large distortions. One method employs a dioptric multi-camera rig system, which reduces and shares the deformation equally over several separate and fixed fish-eye lenses. Further image stitching is required to form an entire panoramic image, which causes the main drawback of this structure that the projection radius should be fixed for a best stitching effect. The Ladybug® system is an example of this case (Sato *et al.*, 2004; Sato and Yokoya, 2010; Ladybug, 2013). The other method uses a linear-array-based camera, which can obtain seamless panoramic images with a vertical and turntable axis, such as the EYESCAN camera system (Schneider and Maas, 2006; Amiri and Gruen, 2010). This structure is not suitable for high-speed platforms, and static or low-speed platforms are preferred (Geyer and Kostas, 2001). A catadioptric system is the third type of panoramic camera composed of several lenses and parabolic mirrors (Geyer and Kostas, 2001; Barreto and Araujo, 2005). This paper concentrates on the multi-camera rig system with spherical projection.

The basic projective geometry of a panoramic camera is still represented by the ideal pinhole model, which describes the co-linearity that 3D object points, corresponding image points in the sphere, and the panoramic center are in a line (see Figure 1a). Kaess and Dellaert used a multi-camera rig for simultaneous localization and mapping (SLAM) with the pinhole spherical sensor model (Kaess and Dellaert, 2010). Paya *et al.* (2010) concentrated on the global description of each omni-directional image. Gutierrez *et al.* (2011) concentrated on the rotation and scale invariance of descriptor patches with a spherical camera model. Spherical perspective transformation functions and stereo-homographies based on the pinhole model are also covered by Mei *et al.* (2008). In Silpa-Anan and Hartley (2005) the fundamental matrix of the pinhole model is used as a geometric constraint between two views. The pinhole model for spherical imaging used in these articles (also referred to in this paper as the ideal panoramic camera model) is adopted under the assumption that the camera contains a unique spherical center as in Figure 1a. In fact, a multi-camera rig system does not contain an entire sphere but consists of several separate lenses with different projection centers and focal lengths (see Figure 1b). This internal structure may introduce additional system errors if the ideal panoramic camera model is applied.

Shunping Ji is with the School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan, 430079, China; and the Center for Spatial Information Science, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8568, Japan.

Yun Shi is with Key Laboratory of Agri-informatics, Ministry of Agriculture / Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, 12 Southern Street of Zhongguanchun, Haidian, Beijing, 10008, China; and the Center for Spatial Information Science, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8568, Japan.(shiyun@caas.cn).

Zhongchao Shi is with the Department of Restoration Ecology and Built Environment, Faculty of Environmental Studies, Tokyo City University, 3-3-1 Ushikubo-nishi, Tuzuki-ku, Yokohama, Kanagawa, 224-8551, Japan.

Anmin Bao and Junli Li are with the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, 818 Beijing South Road, Urumqi, 830011, China.

Xiuxiao Yuan is with the School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road,Wuhan, 430079, China.

Yulin Duan and Ryosuke Shibasaki are with the Center for Spatial Information Science, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8568, Japan.

Photogrammetric Engineering & Remote Sensing
Vol. 80, No. 3, March 2014, pp. 229–238.
0099-1112/14/8003–229

doi: 10.14358/PERS.80.3.229

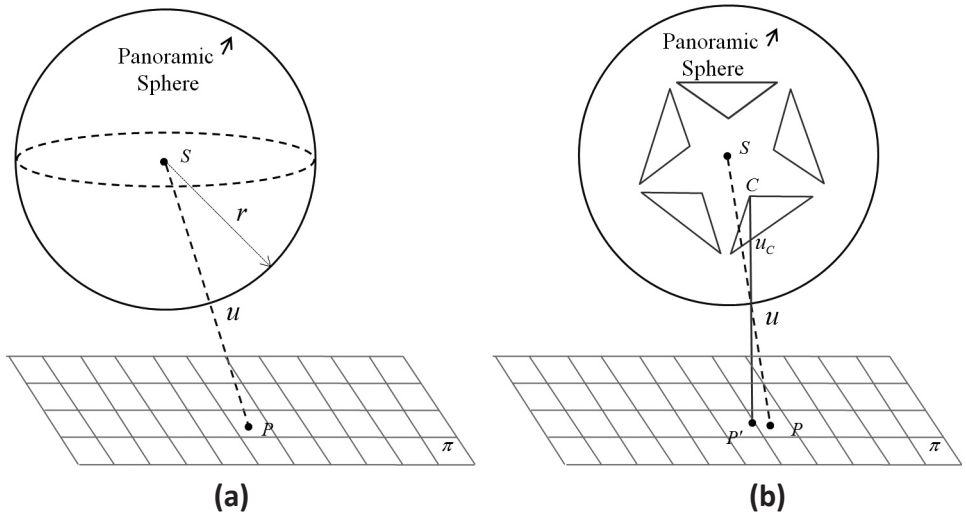


Figure 1. (a) The ideal panoramic camera model shows that the sphere center *S*, 3D points *P*, and panoramic image point *u* are collinear (dashed line); and (b) the rigorous panoramic camera model shows that the mono-camera projection center *C* (solid line), true 3D points *P′* and *u* are collinear. $\boldsymbol{T}_r$ denotes the translation vector from the coordinates of the mono-ideal plane camera to the panoramic coordinates. The difference of *P* and *P′* reflects the localization errors of the ideal camera model.

Generally, two steps are carried out to obtain panoramic images for a multi-camera rig system. First, the separated fish-eye image is rectified to the ideal plane using known calibration parameters. Second, the separate plane images are projected onto a uniform panoramic sphere. Because a fish-eye camera may introduce larger distortions than a plane camera, calibrations should be carried out in advance with adequate geometric constraints, a topic addressed by Kannala and Brandt (2006) and by Fraser (2013) with photogrammetric accuracy. In this paper, we ignore the calibration step and concentrate on analyzing the geometric relationships between a rectified mono-camera, a virtual panoramic camera, and the real 3D points.

In Shi *et al.* (2013), a rigorous camera model that can represent the actual rays (see the solid line in Figure 1b in a multi-camera rig system is proposed. However, in that work, the difference between the two camera models was not compared with real data, and the analytical error distribution of the ideal was not deduced. In this paper, a comprehensive comparison is carried out because application of the ideal model is the most common case. Our contributions can be described as follows. First, the system errors of the ideal panoramic camera model are analytically constructed, including both the in-perspective transformation of a mono-image and the co-planarity of stereo-pair. Second, the applicable conditions of each model for a precise 3D measurement are revealed with a real dataset in different environments.

Sensor Model for a Panoramic Camera

Ideal Panoramic Sensor Model

Figure 1a presents a schematic diagram of the ideal panoramic projection that represents the sphere center *S*, the 3D points *P* in plane π, and the panoramic image point *u* are collinear. To establish the collinear equations, first, the coordinates of point *u* must be obtained from the original mono-images in a multi-camera rig system. As shown in Figure 1b, the panoramic camera is composed of five separate fish-eye lenses. Equation 1 describes how an image point u_c with a coordinate vector $\boldsymbol{X}$ in a separate lens is projected to *u* with a coordinate vector $\boldsymbol{X}$ in the panoramic sphere. $\boldsymbol{K}_r$ represents the transformation matrix from the image coordinate $\boldsymbol{X}$ in the fish-eye camera to the corresponding undistorted plane coordinate, including such calibration parameters as radial distortion, tangential distortion, and the principal point offset (Kannala and Brandt, 2006). $\boldsymbol{R}_r$ and $\boldsymbol{T}_r$ denote the rotation matrix and translation vector from the coordinates of the ideal plane camera to the panoramic coordinates, respectively, and $\boldsymbol{K}_r$, $\boldsymbol{R}_r$, and $\boldsymbol{T}_r$ are fixed values due to the advanced calibration, whereas *m* is the scale factor from the ideal plane to the panoramic sphere coordinate. Equation 2 describes the constraint that all $\boldsymbol{X}' = [x, y, z]^T$ should lie on the surface of a panoramic sphere with a radius *r*. Therefore, the panoramic coordinate $\boldsymbol{X}$ can be solved according to Equations 1 and 2 and is the same value in both the ideal and rigorous model:

$$\boldsymbol{X}' = m\boldsymbol{R}_r\boldsymbol{K}_r\boldsymbol{X} + \boldsymbol{T}_r \tag{1}$$

$$x^2 + y^2 + z^2 = r^2. \tag{2}$$

Next, $\boldsymbol{X}'$ is associated with the real-world coordinate $\boldsymbol{X}_A$ with a perspective transformation in Equation 3, and $\boldsymbol{R}$ and $\boldsymbol{T}$ are the rotation matrix and translation vector between an arbitrary 3D point *P* in the object space and the corresponding panoramic point *u*, respectively, and λ is the scale factor:

$$\boldsymbol{X} = \boldsymbol{R}^T(\boldsymbol{X}_A - \boldsymbol{T}). \tag{3}$$

Figure 1b shows that system errors occur if Equation 3 is used as a pinhole model because it represents a biased ray *Su* (dashed line). However, *Cu*, which passes though the mono-lens center *C* (solid line), represents the correct ray. The biased ray causes localization error because the real 3D point *P′* is translated to an incorrect position *P*.

However, the projection center *C* of the separate fish-eye camera and the panoramic center *S* are quite close, and the angle between *Su* and *Cu* is quite small, which may ensure a relatively small 3D measurement error.

Rigorous Panoramic Sensor Model

In Shi *et al.* (2013), a rigorous sensor model is presented that expresses the correct rays, which means that the co-linearity through *Cu* is established in the uniform panoramic coordinates as shown in Equation 4:

$$\boldsymbol{T}_r + (\boldsymbol{X}' - \boldsymbol{T}_r) = \boldsymbol{R}^{\mathrm{T}}(\boldsymbol{X}_A - \boldsymbol{T}) \; . \tag{4}$$

In Equation 4, $\boldsymbol{T}_r$ is the translation vector between the mono-lens center C and the sphere center S, and $\boldsymbol{X}'$ represents the panoramic coordinate vector, as shown in Equation 1. The vector $\lambda(\boldsymbol{X}' - \boldsymbol{T}_r)$ represents the true ray in the mono-camera coordinate system. After the coordinate origin of the ray is transferred to the panoramic center by adding a translation $\boldsymbol{T}_r$, the rigorous perspective model is constructed by rotation R and translation $\boldsymbol{T}$. We note that the panoramic coordinate $\boldsymbol{X}''$ obtained from Equation 1 should be consistent with $\boldsymbol{T}_r$, which is different for different lenses (see Table 1).

Table 1. Translations From Mono Lenses to Panoramic Center

Lens ID	T_x (m)	T_y (m)	T_z (m)
0	0.0418	–0.0018	–0.0002
1	0.0110	–0.0402	0.0002
2	–0.0349	–0.0229	0.0002
3	–0.0327	0.0260	–0.0002
4	0.0149	0.0389	–0.0000
5	0.0011	0.0011	0.0620

Certain researchers may argue that the system errors caused by the biased rays can be compensated in the prior calibration. For example, if $\boldsymbol{X}'$ is obtained from the intersection of the sphere surface and SP', localization error can be avoided. In fact, it is impossible to obtain an exact $\boldsymbol{X}'$ because the depth of object P' is unknown. Therefore, the compensation values vary with different depths of field. An additional unknown parameter related to depth may be added to $\boldsymbol{X}'$ in Equation 3 as a compensation for the system errors but becomes more complex than Equation 4 and shows not to be a good idea.

Co-planarity of the Two Panoramic Camera Models

The co-planarity between stereo-image pairs reflects the two camera positions and the corresponding image points in one plane. Co-planarity also exists in the panoramic stereo pairs. Figure 2 shows two stereo panoramic images with a panoramic baseline SS'. The dashed lines represent the co-planarity of the ideal camera model and intersect at P; the solid lines represent the co-planarity of the rigorous model and intersect at P'. The baseline SS' is written as $\boldsymbol{B} = [B_X\ B_Y\ B_Z]^{\mathrm{T}}$, ray Su is written as $\boldsymbol{V}_1 = [X_1\ Y_1\ Z_1]^{\mathrm{T}} = [x_1\ y_1\ z_1]^{\mathrm{T}}$, $S'u'$ is written as $\boldsymbol{V}_2 = [X_2\ Y_2\ Z_2]^{\mathrm{T}} = \boldsymbol{R}[x_2\ y_2\ z_2]^{\mathrm{T}}$, and the rotation matrix $\boldsymbol{R}$ is applied between the two images. The vectors $\boldsymbol{B}$, $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$ satisfy the co-planarity:

$$\boldsymbol{B} \bullet (\boldsymbol{V}_1 \times \boldsymbol{V}_2) = 0 \; . \tag{5}$$

If Equation 5 is expanded by the third line of the determinant, the epipolar equation can be deduced from Equation 6, in which a, b, and c are determined by the values of $\boldsymbol{V}_1$ and $\boldsymbol{R}$. Equation 6 represents a 3D plane that passes through the coordinate origin. Combined with Equation 2, it can be concluded that the epipolar line of the ideal panoramic stereo images is a large circle through the projection center. Equation 6 can be used as a geometric constraint for image matching and outlier elimination as well as the traditional plane epipolar constraints:

$$ax_2 + by_2 + cz_2 = 0 \; . \tag{6}$$

Equation 5 is not rigorous because the real rays do not pass through the centers S and S' of the panoramic spheres but rather through the projection centers of the separate lenses C and C'. Thus, the vectors $\boldsymbol{B}$, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ in Equation 5 all contain errors.

In a rigorous camera model, the actual corresponding rays $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ pass through the projection centers of the separate cameras as in Equations 7 and 8, and $\boldsymbol{B}$ is the baseline between the separate lenses but is expressed in the uniform panoramic coordinates, as in Equation 9.

$$\boldsymbol{V}_1 = \boldsymbol{X}_1 - \boldsymbol{t}_1 = \begin{bmatrix} x_1 - T_{x_1} \\ y_1 - T_{y_1} \\ z_1 - T_{z_1} \end{bmatrix} \tag{7}$$

$$\boldsymbol{V}_2 = \boldsymbol{R}(\boldsymbol{X}_2 - \boldsymbol{t}_2) = \boldsymbol{R}\begin{bmatrix} x_2 - T_{x_2} \\ y_2 - T_{y_2} \\ z_2 - T_{z_2} \end{bmatrix} \tag{8}$$

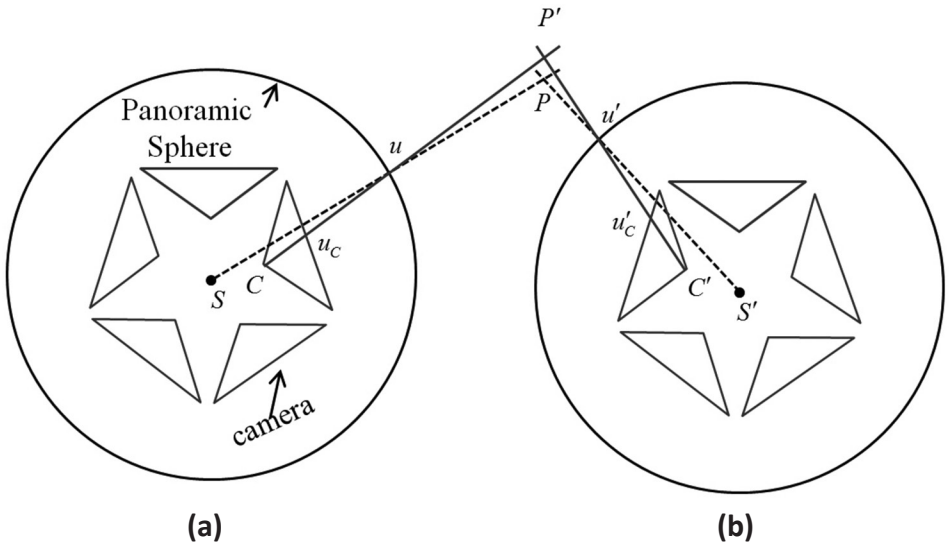


Figure 2. The dashed lines and the baseline SS' represent the co-planarity of the ideal camera model; the solid lines and CC' represent the co-planarity of the rigorous model. $\boldsymbol{R}$ and $\boldsymbol{B}$ are the rotation matrix and translation vector from the panoramic coordinates of the right camera to the reference coordinates of left camera, respectively.

$$\boldsymbol{B} = \boldsymbol{S}_2 - \boldsymbol{S}_1 = \begin{bmatrix} B_X + T'_{x_2} - T_{x_1} \\ B_Y + T'_{y_2} - T_{y_1} \\ B_Z + T'_{z_2} - T_{z_1} \end{bmatrix} \quad (9)$$

In Equations 7, 8, and 9,

$$\boldsymbol{S}_1 = \boldsymbol{T}_1 + \boldsymbol{t}_1 = \begin{bmatrix} X_{S_1} + T_{x_1} \\ Y_{S_1} + T_{y_1} \\ Z_{S_1} + T_{z_1} \end{bmatrix}; \boldsymbol{S}_2 = \boldsymbol{T}_2 + \boldsymbol{t}'_2 = \begin{bmatrix} X_{S_2} + T'_{x_2} \\ Y_{S_2} + T'_{y_2} \\ Z_{S_2} + T'_{z_2} \end{bmatrix}; \boldsymbol{t}'_2 = \begin{bmatrix} T'_{x_2} \\ T'_{y_2} \\ T'_{z_2} \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} T_{x_2} \\ T_{y_2} \\ T_{z_2} \end{bmatrix};$$

$\begin{bmatrix} X_{S_1} \\ Y_{S_1} \\ Z_{S_1} \end{bmatrix}, \begin{bmatrix} T_{x_1} \\ T_{y_1} \\ T_{z_1} \end{bmatrix}, \begin{bmatrix} X_{S_2} \\ Y_{S_2} \\ Z_{S_2} \end{bmatrix}, \begin{bmatrix} T_{x_2} \\ T_{y_2} \\ T_{z_2} \end{bmatrix}$ are the panoramic projection centers and offsets of the two images between the mono-lens and the panoramic camera center, respectively.

If $\boldsymbol{B}$, $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$ are calculated as Equations 7, 8, and 9, Equation 5 will become a rigorous model for stereo panoramic co-planarity. The epipolar line can be also established by expanding Equation 5:

$$ax_2 + by_2 + cz_2 + d. \quad (10)$$

In Equation 10, the constant term d is a small value and is determined by $\boldsymbol{V}_1$, $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, and $\boldsymbol{R}$. Thus, the epipolar line is not a large circle around the panoramic sphere. Generally, d is a rather small value, which makes the epipolar line quite similar to a large circle.

Error Analysis of the Ideal Panoramic Sensor Model

Obviously, Equation 4 represents the rigorous co-linearity, which agrees with the real rays. However, the ideal model (Equation 3) is the most widely used because it is simple and intuitive. If the system errors of the ideal model are trivial and negligible, the ideal model may be a better choice. In this section, a detailed error analysis of the ideal model will be presented to evaluate its impact on the accuracy of the 3D measurements.

Localization Errors in Object Space

Figure 3a represents a panoramic image in the x-z plane with sphere center S and radius r. A mono-lens C is set apart from S with distance t, and SC is its main optical axis. Assuming a real point P' on the plane π, and u as the corresponding point on the panoramic image, κ represents the viewing angle from the optical axis SC, and the object-image distance uP' is represented by d. The ideal panoramic sensor model biases the real point P' to P, and x represents the error. According to similar triangles, $CC'u$, $PP'u$, and $Cu \approx r$ for t are much smaller than r, and Equation 11 can be expressed as:

$$x = \frac{dt \tan\kappa}{r} \quad (11)$$

Equation 11 indicates that the error in the x-axis caused by the ideal panoramic model is proportional to the tangent of the viewing angle and object-image distance, and t and r can be viewed as constants due to the specific multi-camera rig system. Taking the Ladybug®3 system as an example, t is approximately 0.04 m, and r is equal to 20 m for the best stitching effects (Ikeda *et al.*, 2003). The maximum viewing angle can be estimated according to the number of separate mono-lenses. In Ladybug3, five lenses arranged in the horizontal directions separately share a maximum viewing angle of approximately 72°, and thus the maximum κ is 36°. For better imaging effects, d is also restricted. Figure 4a indicates the error changes according to different viewing angles $\kappa \in (0, 36°)$, and object-image distances $d \in$ (-20 m, 16 m) (the directions pointing to the sphere center are positive), which is equal to the depth of field of 4 to 40 m. The medium gray indicates error near 0, whereas the darkest and brightest indicate two extreme values of –0.025 m and 0.02 m, which are obtained when $\kappa = 36°$, $d = -20$ m, and $\kappa = 36°$, $d = 16$ m, respectively. However, in processing procedures such as registration, matching, or bundle adjustment (BA) for panoramic close-range images, only a few features are extracted in the edge region of the mono-lenses, and most of the features may have errors less than 0.01 m. The errors in the y-axis can be also expressed as Equation 11 if κ is taken as the viewing angle in the y-direction.

Space Resection Errors in Image Space

Similarly, the resection error in the panoramic image space of the ideal camera model is represented in Figure 3b and Equation 12. The error changes versus different viewing angles $\kappa \in (0, 36)$ and object-image distances $d \in$ (–20 m, 16 m) are shown in Figure 4b, which indicates a maximum image space error of approximately 0.1 m:

$$x = \frac{dt \sin\kappa}{d + r} \quad (12)$$

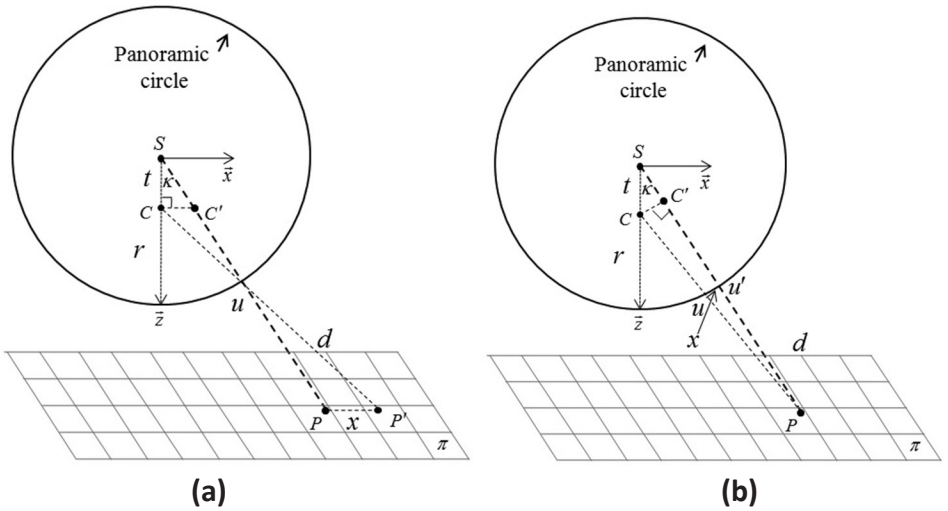


Figure 3. (a) The localization error in the x-direction caused by the ideal camera model; (b) the space resection error in the panoramic image space.

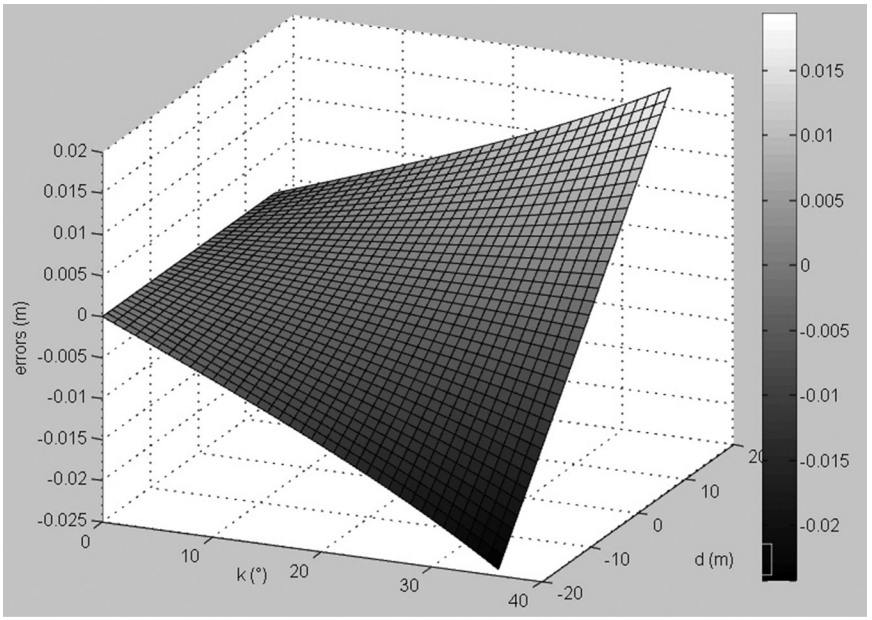
Epipolar Errors

For the stereo pairs, epipolar errors exist in the ideal model compared with the rigorous equations, as shown in Equations 7 through 10. The exact expressions of a, b, c, and d are complex and not intuitive, and thus a simplified and general situation is presumed to reveal the error distribution of the ideal epipolar geometry. First, only horizontal mono-lenses are considered, meaning that the translation of the z-direction from all lenses to the sphere center is equal to 0, which is supported by the real values in Table 1. Second, the angles between the two images are not overly large and the corresponding points are located in the same mono-lens, which guarantees $\boldsymbol{B} \approx [B_X\ B_Y\ B_Z]^{\mathrm{T}}$ in Equation 9. Next, Equation 13 is obtained by substituting Equations 7 through 9 into Equation 5, which is then expanded in Equation 14 according to the distributive law. The left side of Equation 14 represents the ideal co-planarity of the panoramic camera, and the value in the right side unequal to 0 reflects the epipolar bias between the ideal model and the rigorous model. The right side is subsequently expanded in Equation 15, where the small value $\boldsymbol{t}_1 \times \boldsymbol{t}_2'$ is ignored, $B_Z = 0$, and $Z_1 = Z_2$ for a plane shooting.

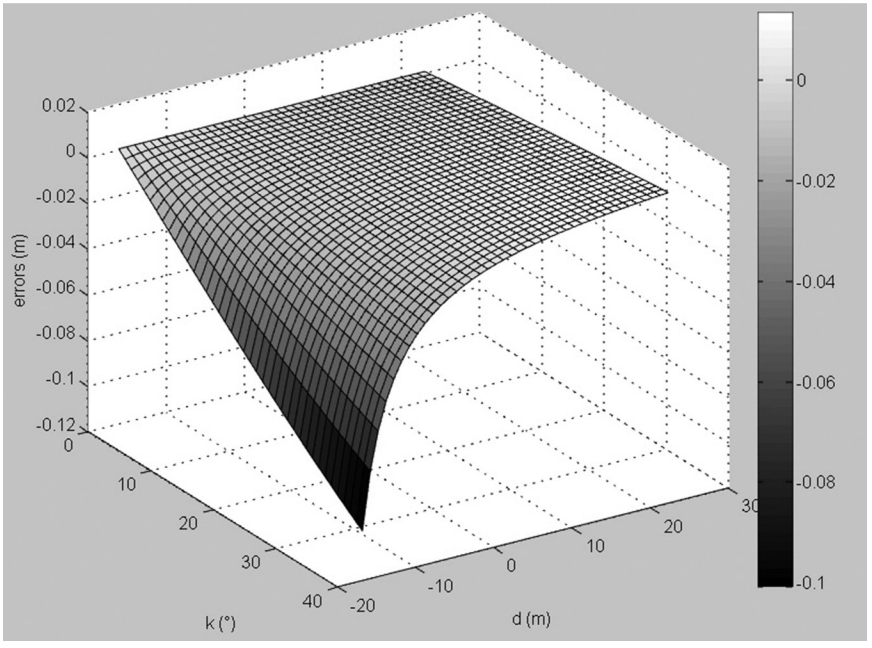
$$B \cdot ((X_1 - t_1)(X_2 - \boldsymbol{t}_2')) = 0 \tag{13}$$

$$B \cdot (X_1 \times X_2) = B \cdot (X_2 \times \boldsymbol{t}_2' + X_2 \times \boldsymbol{t}_1 - \boldsymbol{t}_1 \times \boldsymbol{t}_2') \tag{14}$$

$$B \cdot (X_1 \times X_2) \approx \begin{bmatrix} B_X & B_Y & B_Z \\ X_1 & Y_1 & Z_1 \\ T'_{x_2} & T'_{y_2} & T'_{z_2} \end{bmatrix} + \begin{bmatrix} B_X & B_Y & B_Z \\ X_2 & Y_2 & Z_2 \\ T_{x_1} & T_{y_1} & T_{z_1} \end{bmatrix} \approx Z_1 (B_Y (T'_{x_2} - T_{x_1}) - B_X (T'_{y_2} - T_{y_1})) \tag{15}$$



(a)



(b)

Figure 4. (a) The localization errors in the x-axis change with the viewing angle κ and the object-image distance d in object space; and (b) the space resection errors in the panoramic image space.

In Equation 15, the error of the ideal panoramic epipolar is related to the rotation matrix $\boldsymbol{R}$ and z-coordinate of the reference image point. If there is no rotation between the two images, $\boldsymbol{t}_2'$ equals $\boldsymbol{t}_1$, and the epipolar error is equal to 0. In contrast, the epipolar error increases when the rotation or Z_1 increases. The B_X represents the scale of the stereo pairs and can be set to 1, and in most cases, B_Y is smaller than B_X but may be larger in certain close-range shooting situations. Figure 5 shows the changes in the epipolar errors versus Z_1 and κ, where κ is the rotation angle of the z-axis and is set to (0, 30°).The other two angles of the x-axis and y-axis are set to 0 for plane rotation, Z_1 is set to (–16 m, 16 m), and B_Y is set to 0.2 m. A gradually increasing error occurs when κ and the absolute value of Z_1 increase. An extreme of 0.25 m occurs at $Z_1 = \pm 16$m, $\kappa = 30°$, which represents a relatively larger error than the localization error of a single ray as obtained in Figure 5. In fact, this error is calculated on the sphere with a 20 m radius. The equivalent pixel size at a distance d from the camera can be calculated as in Equation 16. The focus of the mono-camera f is equal to 3.3 mm, and the CCD size p_0 is approximately 4.4 μm in the Ladybug3 system. Therefore, the corresponding pixel size is approximately 0.026 m at a distance of 20 m, which means that the maximum error of the epipolar is approximately 10 pixels, which is a relatively small value.

$$pixel = \frac{d}{f} p_0 \tag{16}$$

Experiments and Results

Test Design

To verify the error in the ideal model and its impact on the 3D measurements, two experiments are conducted in real environments. The first test is carried out in outdoor environments using a car-mounted Ladybug3 camera, and three images are taken with a baseline of 1 m in an area with precise ground control points (GCP). Figure 6a shows a panoramic image of a campus. Eight GCPs are measured using a static GPS with 2 cm accuracy. The second test is carried out in an indoor environment. Five images are taken in a narrow corridor outside of our laboratory by a Ladybug3 camera mounted on a handcart, as shown in Figure 6b. Figure 6b presents a panoramic view of a point-cloud image obtained by a high-precision lidar scanner with approximately 1 mm accuracy. The ten point marks are selected as the GCPs. The indoor and outdoor tests are quite different. The indoor images are taken with a depth of field of approximately 3 to 8 m, and the rotation angles are relatively larger than those of the outdoor images with an average angle of 15°. In contrast, the outdoor images are taken at a straight road with no rotation and a depth of field of approximately 10 to 25 m. The depth of field and rotation angle are the key parameters for the errors of the ideal sensor model as described in Equation 11 and Equation 15.

Table 1 shows the translations from six mono-lenses to the panoramic center. The translations of lenses No.0 to 4 are nearly equal to 0 in the z-axis and nearly equal to 0.04 m in the x-y plane. The No.5 lens points to the sky.

The first comparison is carried out between space resections of the two models with the GCPs, and the residual errors

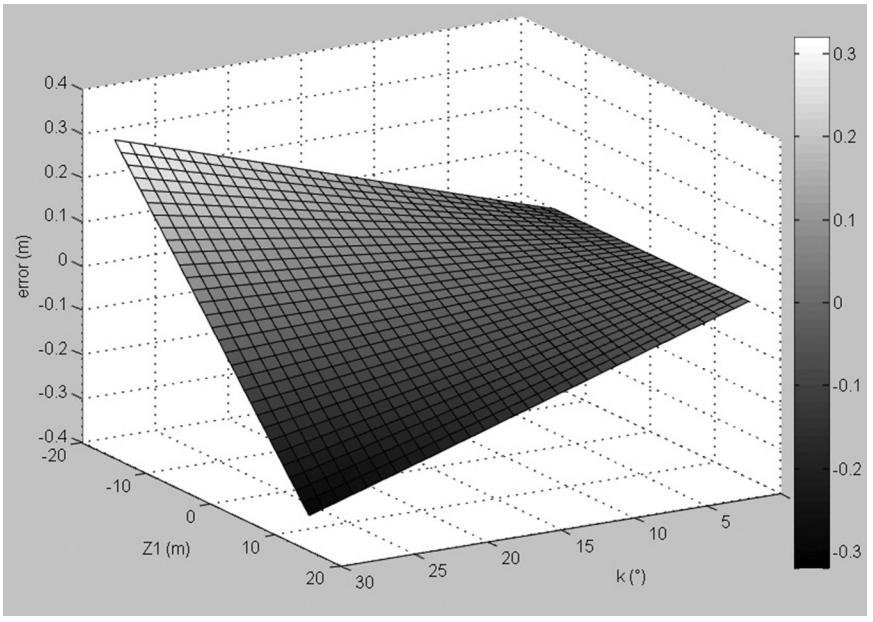


Figure 5. The epipolar errors changes with the rotation angle κ and z-coordinate of reference image Z_1.

Table 2. Space Resection Results of the Two Models

Models		GCP Number	σ_0 (m)	Accuracy of EO	
				Translation(m)	Angle(rad)
Ideal	Outdoor	8	0.0201	0.003	$2.5*10^{-5}$
Rigorous			0.0189	0.003	$2.8*10^{-5}$
Ideal	Indoor	10(11)	0.046(0.120)	0.004	$3.3*10^{-5}$
Rigorous			0.026(0.028)	0.002	$1.7*10^{-5}$

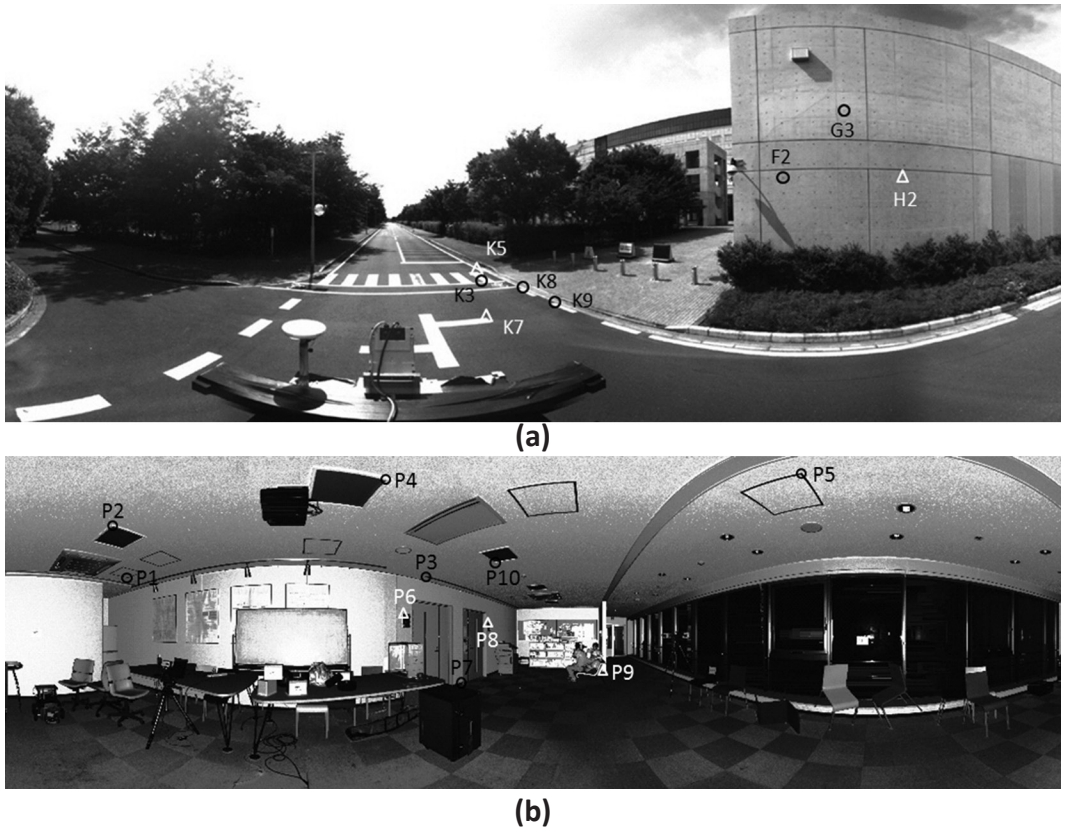


Figure 6. (a) Panoramic image taken outdoors by the Ladybug3 system with eight GCPs; and (b) a point cloud image in a narrow corridor taken by a high-precision lidar scanner from which ten 3D points are selected as the GCPs. The white triangles represent true control points, and black circles represent check points.

are analyzed in image space. The second comparison is designed to analyze the epipolar errors. The last comparison is carried out between a common block triangulation of the two models, and the 3D localization errors are emphasized.

Space Resection Results and Analysis

Table 2 shows the result of space resection. In the outdoor experiments, the root mean square error (σ_0) and the accuracy of the exterior orientation (EO) are nearly the same from both models. Figure 4b shows that the 3D location error of the ideal model is affected by the joint action of the viewing angle and the object distance. Therefore, although several GCPs may possess a larger viewing angle, a depth of field between 10 to 20 m for all of the GCPs excludes large errors.

Obvious differences appear in the indoor test. The σ_0 of the ideal model is 2 cm larger than that of the rigorous model, and the accuracy of EO is relatively lower. The GCPs with large viewing angles produce larger errors than those in the outdoor test, mainly due to the depth of field of approximately 3 to 8 m. The contents in brackets represent the changes of σ_0 in the two models when one GCP located at the No.5 lens (top lens) is added. The σ_0 shows almost no change using the rigorous model but reaches as high as 12 cm when using the ideal model, a large error that is equivalent to greater than four pixels; this indicates that larger errors will be introduced if the No.5 lens is used in the ideal model, primarily because this lens is not located in the same horizontal plane as the others (see Table 1).

Epipolar Error Analysis

To examine the epipolar errors, the precise tie-points between stereo images are first extracted to calculate the relative orientation elements. Next, the left image points and the relative orientation elements are used to calculate the epipolar lines according to Equation 6 and Equation 10, respectively. The minimum distance between the corresponding points in the right image and the corresponding epipolar lines represents the epipolar errors.

Figure 7 presents the epipolar errors of 79 correspondents with a heading angle of 6.88°. The other two angles are nearly 0 for a plane shooting. The epipolar errors of the rigorous model have an average of 0.47 pixels (dotted line) and agree with the image matching errors. In addition, the epipolar errors tend to slightly increase as the z-coordinate values increase due to tiny calibration errors about 2 mm in Table 1. If the ideal model is used, the epipolar errors are larger, with an average of 1.88 pixels, and show an obvious proportional increase with the values of the *z*-coordinates with a fixed rotation and agree with the theoretical error distribution represented by the solid line according to Equation 15.

BA Results and Analysis

Table 3 shows a general comparison of the BA results between the two models. In the outdoor test, the root mean square error (RMSE), accuracy of exterior orientation (EO) and the average 3D intersection error of corresponding rays are nearly the same. The rigorous model results may be slightly better. The RMSE of the panoramic images is approximately 0.02 m, much larger compared with the RMSE of a traditional plane camera with several micrometers. However, the pixel size of the panoramic images is 0.026 m at a distance of 20 m as obtained from Equation 16, which shows that the RMSE is equal to less than 1 pixel and agrees with the measurement accuracy. Another characteristic is that the RMSE is almost equal to the average 3D intersection error of all tie-points in object space. This result can be explained by the observation that the average depth of field is close to 20 m, which is the radial length of the panoramic sphere. Figure 8a displays the check errors of the GCPs. Points K7, K5, and H2 are used as control points, and the

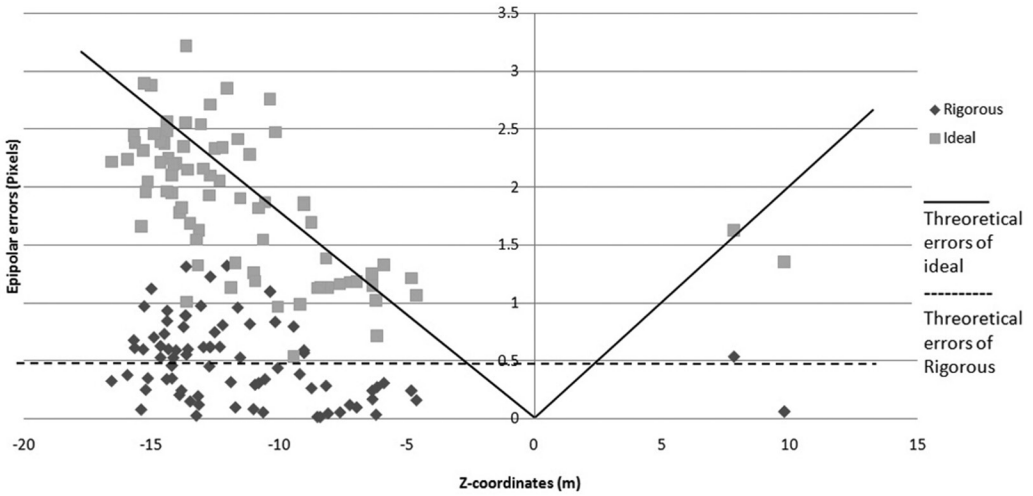


Figure 7. Comparison of the epipolar errors from the two models in the indoor test. The solid line indicates the theoretical errors of the ideal model according to Equation 15, and the dashed line indicates the average epipolar error of approximately 0.47 pixels.

TABLE 3. BA RESULTS OF THE TWO MODELS

Models		σ_0 (m)	Accuracy of EO		3D intersection error(m)
			Translation(m)	Angle(rad)	
Ideal	Outdoor	0.0212	0.012	$3.5*10^{-5}$	0.0214
Rigorous		0.0198	0.012	$3.7*10^{-5}$	0.0200
Ideal	Indoor	0.0401	0.017	$7.3*10^{-5}$	0.0412
Rigorous		0.0210	0.011	$4.5*10^{-5}$	0.0226

remaining points are used as checkpoints. The average errors of the checkpoints are approximately 0.0568 m in the ideal camera model and 0.0526 m in the rigorous model. This difference is rather slight, and the accuracy of the rigorous model is approximately 4 mm higher than the ideal, which can be ignored because the accuracy of the GCPs is only 2 cm.

According to Figure 8a and Table 3, we can conclude that the difference between the ideal and the rigorous model is quite small and can be ignored in this test. This conclusion agrees with our theoretical analysis, as shown in Equation 11. In the outdoor environments of our test, the object-image distance d is between –10 m and 5 m in most cases, which proves that the maximum error is no greater than 1 cm. As a result of adjustment of all tie-points, the errors are merely several millimeters.

Table 3 also shows a basic comparison of the indoor tests between the two models and shows larger differences in all of the reference indices between the ideal and the rigorous model than those of the outdoor test. The RMSE and 3D intersection error of the ideal model are approximately 0.04 cm, which is twice that of the rigorous model. Figure 8b shows the 3D localization errors of the checkpoints. Points P6, P8, and P9 are used as control points, and the remaining points are used for error check. The average error of seven check points is 0.0268 m with the rigorous model and 0.0526 m with the ideal.

The indoor test shows that the differences between the ideal and rigorous panoramic camera model are too remarkable to be ignored. The object distance is approximately 3 to 8 m, which implies an object-image distance of –17 m to –12 m. In this case, the situation may introduce errors of almost 2 cm, according to Equation 11 with an average viewing angle of 15°. This explains why the location errors of the ideal model are approximately 2 cm larger than that of the rigorous model.

Compared with the obvious difference of the ideal model in two tests, Table 2 shows that the performance of the rigorous model in the indoor and outdoor tests is nearly the same and that only the accuracy of the checkpoints is different. The indoor check of the accuracy of the rigorous model is approximately 0.027 m and matches the intersection error, which is approximately 0.02 m. Because the accuracy of the GPS is merely 0.02 m, the outdoor accuracy is approximately 0.05 m. If GCPs with higher accuracy are used, the check error might match the intersection error.

More Experiments

The two experiments have demonstrated the localization accuracy of the ideal model is mainly affected by the object-image distance. Another factor, rotation angles, which affects epipolar, has no effect on localization according to Equation 11. However, Equation 11 only expresses the localization errors of a single ray, while in triangulation of multi-rays, the rotation angles may introduce system errors. Two more experiments are presented to demonstrate that whether rotation angles are an important factor on localization accuracy. The third test was carried out in outdoor environments using the same car-mounted Ladybug3 camera, but eight images were taken in one corner with rotation angle of 90°, against 0° in the first test. Figure 9 shows two images with different sight angles where nine GCPs are marked. The fourth test was performed in the same indoor environment as shown in Figure 6b, but the rotation angles of three taken images are almost 0°, against 15° in the second test.

Figure 10a demonstrates the localization errors of the two different models in the third test. The difference between the ideal and the rigorous model is small enough to be ignored. While in the fourth test, Figure 10b demonstrates that the location errors of the ideal model are approximately 2 cm larger than that of the rigorous model, the same as that in the second test. From all the tests, we can conclude that the rotation

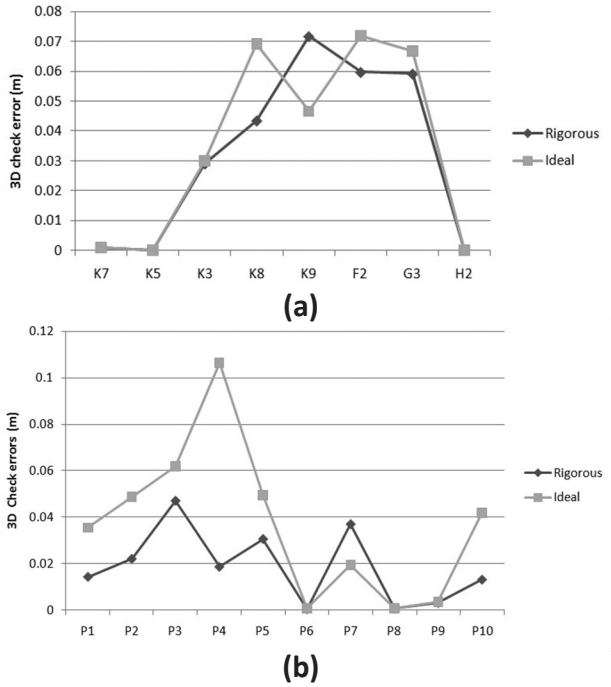


Figure 8. (a) A 3D error check comparison of the two models in the outdoor test; and (b) a 3D error check comparison of the two models in the indoor test.

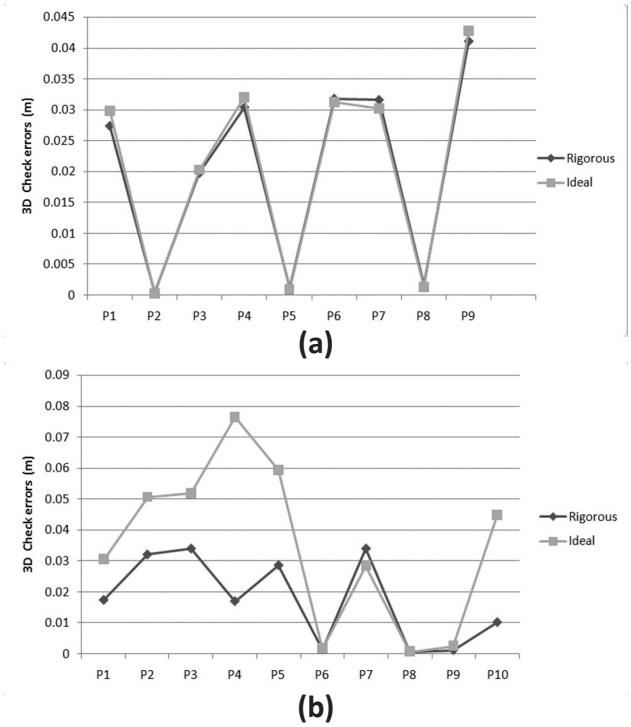


Figure 10. A 3D error check comparison of the two models: (a) the third test in the outdoor environment with 90° rotation angles; and (b) the fourth test in the indoor environment with 0° rotation angles.

Figure 9. Images taken at the corner by the Ladybug3 system with nine GCPs, in which three white triangles represent control points and six black circles represent check points: (a) part image from lens No.0 (front view); and (b) from lens No.3 (back-side view).

angles is not a considerable factor for localization accuracy, because the ideal model in the two outdoor tests achieved the same accuracy as the rigorous model and the same poorer accuracy against the rigorous model in the two indoor tests, although with different rotation angles, respectively.

Another factor, viewing angles, have more complex impacts on localization than object-image distance and rotation angles. The impacts are simple for single ray as in Equation 11, while in triangulation of multi-rays with different viewing angles, predicting the right localization errors is quite difficult. For example, the errors produced by two correspondent rays with large viewing angles may compensate each other. However, viewing angles are not so important and are subject to object-image distance. From the two outdoor tests (Figure 7a and Figure 10a), we can see the localization accuracy of all check points are normally distributed in spite of large viewing angle differences. That means the viewing angles have no effect on localization accuracy conditioned on small object-image distance. The indoor test (Figure 8a) shows the check point P4 has larger errors than the others due to large sight angles about 30° in *y*-direction. However, the average localization errors of all checkpoints still agree with our theoretical analysis, i.e., 2 cm system errors are introduced by the ideal model with an average viewing angle of 15° at 15 m object-image distance. So, considering well distributed tie points or GCPs, we can easily use the average viewing angle for error estimation, conditioned on different object-image distances. Object-image distance is still the most important factor.

Conclusions

In this paper, a widely used ideal panoramic camera model for a multi-camera rig is carefully analyzed and compared with a rigorous model using precise 3D measurements in different environments. To the best of our knowledge, the rigorous model only appears here and in (Shi *et al.*, 2013). The outdoor test shows that the difference between the two models is notably slight, and both models are applicable for vehicle localization and 3D measurements in the environment where the average object-image distance is less than 10 m. Please note that it is not limited to outdoor locations, and some indoor environments such as a spacious hall can meet the requirement. In contrast, the indoor test with an average object-image distance about 15 m shows that the rigorous model is stricter and produces better measurement accuracy than the ideal model. The difference between the two test results supports our theoretical analyses of the system errors caused by the ideal model. In other words, the errors increase when the distance between the objects and the panoramic sphere increase and the viewing angles increase. Another conclusion is that the epipolar errors of the ideal camera model are influenced by the *z*-coordinates and rotation angles, which may mean that correct tie-points with larger *z*-coordinates and larger rotations are more likely to be removed in an error excluding procedure.

We recommend that the rigorous panoramic sensor model is used due to its rigor in geometry, especially under the environments where objects are too near or too far from the panoramic sphere, in other words, a large object-image distance exists. The ideal model contains another drawback. If a large rotation occurs (e.g., a correspondence appears in the No.0 and No.5 lenses), more errors will be introduced than in our theoretical analysis above. Another advantage of the rigorous model is that it is independent of the projection surface. For example, if a cylinder projection is used, Equation 1 and Equation 4 remain the same, and only the sphere constraints (Equation 2) should be changed to cylinder constraint equations. In contrast, the system errors of the ideal model are relevant to the different projection surfaces and should be analyzed separately.

However, the ideal model may still be effective in certain environments or applications. First, under the environments where object-image distance is less than 10 m, the 3D measurements with the ideal model are nearly the same as those from the rigorous model. Second and most importantly, many localization applications may not require photogrammetric accuracy up to several centimeters. For example, a long-distance real-time localization effort (Tardif *et al.*, 2008; Zheng *et al.*, 2013), whether indoor or outdoor, may exhibit accumulative localization errors of more than several meters. So errors of several centimeters caused by the ideal model can clearly be ignored.

Acknowledgments

This work was supported by the Chinese 973 Program (2012CB719902), the GRENE project of MEXT Japan, the National Natural Science Foundation of China (41301365), the International Science & Technology Cooperation Program of China (2010DFA92720-04), the opening project of the Key Laboratory of Xinjiang Uygur Autonomous Region (XJYS0205-2012-02), and the Key Program of the Chinese Academy of Sciences (KZZE-EW-08-02-02).

References

Amiri, P.J., and A. Gruen, 2010. Sensor modeling, self-calibration and accuracy testing of panoramic cameras and laser scanners, *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):60–76.

Anguelov, D., C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, 2010. Google street view: Capturing the world at street level, *Computer*, 43(6):32–38.

Fraser, C.S., 2013. Automatic camera calibration in close range photogrammetry, *Photogrammetric Engineering & Remote Sensing*, 79(4):381–388.

Gutierrez, D., A. Rituerto, J.M. Montiel, and J.J. Guerrero, 2011. Adapting a real-time monocular visual SLAM from conventional to omnidirectional cameras, *Proceedings of the 11th OMNIVIS at the IEEE International Conference on Computer Vision (ICCV)*, 06-13 November, Barcelona, Spain, pp. 343–350.

Ikeda, S., T. Sato, and N. Yokoya, 2003. High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system, *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems(MFI2003)*, 30 July-01 August, Tokyo, Japan, pp. 155–160.

Kaess, M., and F. Dellaert, 2010. Probabilistic structure matching for visual SLAM with a multi-camera rig, *Computer Vision and Image Understanding*, 114(2):286–296.

Kannala, J., and S.S. Brandt, 2006. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340.

Ladybug, 2013. URL: *http://www.ptgrey.com/products/spherical.asp* (last date accessed: 17 December 2013).

Li, R., K. Di, L.H. Matthies, R.E. Arvidson, W.M. Folkner, and B.A. Archinal, 2004. Rover localization and landing site mapping technology for the 2003 Mars Exploration Rover Mission, *Photogrammetric Engineering & Remote Sensing*, 70(1):77–90.

Mei, C., S. Benhimane, E. Malis, A. Member, and P. Rives, 2008. Efficient homography-based tracking and 3-D reconstruction for single-viewpoint sensors, *IEEE Transactions on Robotics*, 24(6):1352–1364.

Paya, L., L. Fernandez, A. Gil, and O. Reinoso, 2010. Map building and Monte Carlo localization using global appearance of omnidirectional images, *Sensors*, 10(12):11468–11497.

Sato, T., S. Ikeda, and N. Yokoya, 2004. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system, *Camera*, 2:326–340.

Sato, T., and N. Yokoya, 2010. Efficient hundreds-baseline stereo by counting interest points for moving omni-directional multi-camera system, *Journal of Visual Communication and Image Representation*, 21(5-6):416–426.

Schneider, D., and H. Maas, 2006. A geometric model for linear-array-based terrestrial panoramic cameras, *The Photogrammetric Record*, 21(115):198–210.

Shi, Y., S. Ji, Z. Shi, Y. Duan, and R. Shibasaki, 2013. GPS-supported visual SLAM with a rigorous sensor model for a panoramic camera in outdoor environments, *Sensors*, 13(1):119–136.

Silpa-Anan, C., and R. Hartley, 2005. Visual localization and loopback detection with a high resolution omnidirectional camera, *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks*, 21 October, Boston, Massechuttes, unpaginated CD-ROM.

Tardif, J.P., Y. Pavlidis, and K. Daniilidis, 2008. Monocular visual odometry in urban environments using an omnidirectional camera, *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, 22-26 September, Nice, France, pp. 2531–2538.

Zheng, B., Y. Dong, A. Davies, B. Mullany, and E. Morse, 2013. Using optical projection in close-range photogrammetry for 6DOF sensor positioning, *Photogrammetric Engineering & Remote Sensing*, 79(1):79–86.

(Received 02 June 2013; accepted 25 August 2013; final version 26 September 2013)