Journal of Universal Computer Science, vol. 16, no. 10 (2010), 1269-1290
submitted: 30/6/08, accepted: 9/3/10, appeared: 28/5/10 © *J.UCS*

Pose Estimation of Rotating Sensors in the Context of Accurate 3D Scene Modeling

Karsten Scheibe
(German Aerospace Center (DLR), Berlin, Germany
karsten.scheibe@dlr.de)

Fay Huang
(CSIE, National Ilan University, Yi-Lan, Taiwan
fay@niu.edu.tw)

Reinhard Klette
(Department of Computer Science, The University of Auckland, New Zealand
r.klette@auckland.ac.nz)

Abstract: Sensor-line cameras have been designed for space missions in the 1980*s*, and are used for various tasks, including panoramic imaging. Laser range-finders are able to generate dense depth maps (of isolated surface points). Panoramic sensor-line cameras and laser range-finders may both be implemented as rotating sensors, and we used them together this way to reconstruct accurately 3D environments (such as, for example, large buildings).

This article reviews related developments, followed by a detailed description of designed calibration and pose estimation techniques which have been used for both rotating sensors. Related experiments evaluate the accuracy of calibrated sensor parameters and of estimated poses.

Key Words: panoramic imaging, camera calibration, pose estimation, sensor-line camera

Category: I.3.5, I.4.8

1 Introduction

Panoramic images can be generated by various techniques, such as mosaicing or stitching [Chen 1995], or can also be acquired using specialized sensors such as catadioptric [Nayar 1997] or rotating cameras [Reulke and Scheele 1998]. Many applications in computer vision, computer graphics (e.g., image-based rendering) or photogrammetry, demand spatial (geometric) or color accuracy, very high resolution, or minimized radial distortion, which excludes the use of simple stitching, mosaicing or catadioptric camera. For example, the quality of close-range photogrammetry (e.g., for static scenes of architecture) depends on image resolution to support very accurate representations of scene geometry. Computer animations in a professional context often specify needs for very high resolution, photometric correctness, or geometric accuracy. The book [Huang et al. 2008] deals with panoramic sensors and their applications for the generation of (very) high resolution 3D models; this paper describes the sensor calibration techniques as discussed in this book but in more detail, with further experiments and thus also with more concluding comments.

The scan geometry of rotating laser range-finder (LRF) and rotating sensor-line camera is very similar, and this supports accurate rendering of 3D surfaces, generated from range-scans, using color panoramic images, recorded with a rotating sensor-line camera. The fusion problem of panoramic image data and LRF depth data has also been discussed in [Jiang and Lu 2007], in which an imaging model was used which combines a regular digital camera with an LRF on a turntable. See also [Parian and Gruen 2004] for an alternative model of a rotating sensor-line camera.

In this article, we put special emphasis on "high-resolution" panoramic images taken by a rotating sensor-line camera. The two sets of data (i.e., panoramic image and laser scan) used for fusion can be obtained separately without special spatial constraint. Actually, multiple panoramic images are generally required to support texture information as some surface patches may only be seen by one panoramic camera but not another. Thus, extrinsic parameter calibration is of crucial importance, and we discuss our calibration in detail in this paper. To the best of our knowledge, no one else has tackled the pose estimation problem from multiple rotating sensor-line cameras so far.

1.1 Panoramic Sensors

Informally speaking, a panoramic sensor sends rays into the 3D world. A ray emerges at a *projection center* and collects information (such as color, range, or intensity) about the first surface point it hits in 3D space (possibly also involving recursion as known from ray tracing in computer graphics).

In an abstract geometric sense, measured data are mapped onto a *capturing surface*, which may be understood as being a plane in the case of a "normal camera" (note: a CCD[1] or CMOS[2] matrix of sensor elements defines basically a photosensitive rectangle), or a sphere, cube, cylinder, and so forth in the case of a "non-standard panoramic sensor" (and this defines *spheric*, *cubic*, or *cylindric panoramas*). With respect to the capturing surface, catadioptric panoramas may be called *hyperboloidal panoramas*. See [Daniilides and Klette 2006] for various architectures of panoramic sensors and image geometries.

Extreme cases of panoramas are defined by a full $360° \times 360°$ *spherical view* or *cubic view* (see, for example, Google's street view), a full $360°$ *cylindrical view*

[1] Charged coupled devices.
[2] Complementary metal oxide semiconductor.

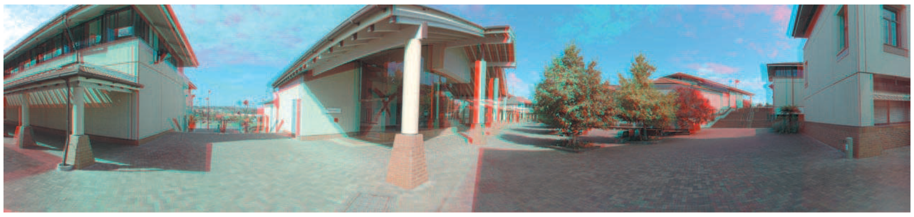

Figure 1: Anaglyphic panorama of buildings at Tamaki campus, The University of Auckland.

(e.g., as used for panoramic X-rays in dentistry), or a *translational planar view* (an example of an *image mosaic*). A panorama can also be defined by a subrange of one of those listed examples, as long as it still allows a wider view compared to normal photographs. We may still call it a cylindric or spheric panorama. The image in Figure 1 is a segment of a cylindric anaglyphic panorama (generated by the authors) which requires anaglyphic eyeglasses for proper stereo viewing. High-quality and high-resolution digital cameras and fish-eye lenses are also increasingly alternative options for generating high-quality panoramic images.

1.2 Rotating Sensor-Line Cameras

Imagine that the sensor matrix of a "normal digital camera", consisting of $M \times N$ sensor elements (each recording a single pixel), degenerates in a way that there is only one column of sensor elements (i.e., $N = 1$; for example, similar to those used in a flatbed scanner). The benefit is that current sensor technology allows to produce such a *sensor-line* for very large values of M, say M greater than 10,000, but producing sensor-matrices of 10,000 $\times$ 10,000 elements at decent costs is still a challenge today.

A digital camera, with the sensor-matrix "shrunk" into a single sensor-line, may now be placed on a tripod and rotated, taking many images, "column by column" during such a rotation. This defines a *rotating sensor-line camera*, a panoramic sensor which may record 360° panoramic images within a time frame needed for taking many shots during one full rotation. Such a sensor is not only more economic (compared to the use of a, say, 10,000 $\times$ 10,000 sensor-matrix camera), it also comes with several benefits for recording panoramic images.

In the 1990s, theoretical studies by various authors (e.g., [Ishiguro et al. 1992]) pointed out that the use of a rotating sensor-line camera, where panoramas are shot line by line, each line with its own projective center, allows to control conditions for improved stereo analysis and stereo viewing. Basically, this was the start into a new category of digital panoramas, defined by super-high resolution and

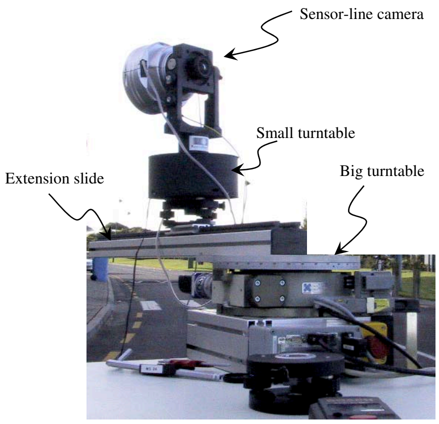

Figure 2: An experimental rotating sensor-line camera configuration using an Eyescan M3 and an additional extension slider.

geometric accuracy. Figure 2 shows the experimental rotating sensor-line camera configuration of the authors using an Eyescan M3. Many of our experimental panoramic images including the stereo panorama in Figure 1 were captured by using this sensor system.

Sensor-line cameras actually had been designed for digital aerial imaging (using a *push-broom technique*) already since the early 1980s. Three or more of such sensor lines, optics, and a frame grabber together define a *sensor-line camera* to be used in todays airborne image sensors (e.g., for creating "3D maps" of cities or interesting landscapes, see [3D Reality Maps^{TM}]).

These physically existing sensor-line cameras were positioned and rotated on tripods, thus generating panoramas. For example, the *Wide Angle Airborne Camera* (WAAC) of DLR[3] Berlin-Adlershof, see [Reulke and Scheele 1998], was used in 1995 for taking the first cylindric panoramic image (a view from the roof of Dornier in Germany) using a rotating sensor-line camera. The panoramic image is of very high resolution for that time, with each of its columns consisting of more than 5,000 pixels.

Panoramic images of very high resolution are today required in various applications, such as inspections of pipelines (as an example of industry applications in general), scanning of house facades for 3D city maps, or accurate indoor and outdoor documentation of selected architectural sites. In the latter two cases, panoramic images are used for texture mapping, and the 3D geometry of the ob-

[3] The German Air-and-Space Institute (DLR = Deutsche Luft- und Raumfahrt) has actually various institutes in Germany, with its head institute near Munich.

Figure 3: Use of a laser range-finder for generating 3D models of buildings on Tamaki campus, The University of Auckland.

ject surfaces is typically acquired by a laser range-finder (on a moving platform, or at multiple locations).

1.3 Laser Range-Finder

An LRF or *laser scanner* determines distances to opaque objects; it is also known as LIDAR (Laser Imaging Detection and Ranging). It records distances (at accurately recorded horizontal and vertical angular increments) between projection center and surface points which generates a *range-scan*. The particular case of returning intensity only (rather than color) forms a *gray-level image*. A range-finder typically allows to obtain such a gray-level image in addition to its range-scan (but not a color image); a pair consisting of a range-scan and a gray-level image is basically geometrically aligned because sensed by the same sensor at the same viewpoint.

An LRF determines the distance to an object or surface using laser pulses (similar to radar technology, which uses radio waves instead of light). Each individual pulse is directed along one scan ray, and the returned range value identifies one point in 3D space. Figure 3 illustrates the use of such a device: for this model of an LRF, the laser pulses radiate through the rectangular window, and the small circular window covers a sensor-matrix camera for capturing (relatively low-resolution) color images, sufficient for identifying locations. The figure shows on the right such a color image and a visualization of a range-scan in image form. Generated models of buildings allow us, for example, to generate "ground truth" for the performance evaluation of stereo and motion analysis in vision-based driver assistance [EISATS].

A produced range-scan actually defines a "cloud" of points in 3D space, which represents visible surfaces by those (possibly noisy) discrete points. These

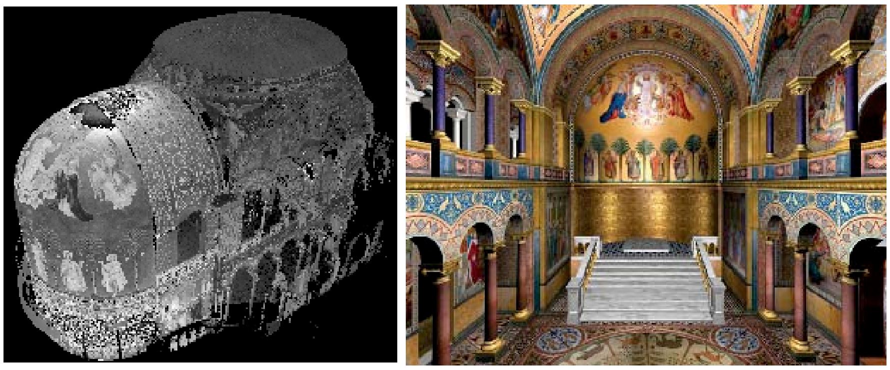

Figure 4: Left: "cloud of points" illustrating multiple laser range-scans of throne room of Neuschwanstein castle. Right: a view into the reconstructed 3D model with color texture generated by multiple high-resolution panoramic scans.

isolated points need to be mapped into meshed (e.g., triangulated) surfaces, and the surfaces may be "smoothly" rendered using gray-levels recorded by the range-finder. However, color is typically requested for rendering, and color panoramic images may be used for proper rendering. Figure 4 illustrates on the left the data obtained by LRF scans, and a view of the high-resolution 3D reconstruction result on the right, calculated by using multiple LRF scans and panoramic images.

2 Sensor Geometry

We consider two sensors of very similar architecture and unify their description by considering both (i.e., also the LRF) as being panoramic sensors.

2.1 Panoramic Sensors

The projection center of the sensor-line is denoted as $\mathbf{C}_i$ (for $i \in \mathbb{N}$), which describes the position of the sensor-line camera. As the camera is rotated 360 degrees along a pre-specified axis, the trajectory of the camera projection center defines a circle called *base circle* (see Figure 5). Ideally, we assume that the plane of the base circle is perpendicular to the rotation axis, the camera's optical axis remains coplanar to the base circle at all of its positions during the rotation, and the sensor-cell array is configured parallel to the rotation axis.

Through such a 360°-rotation, the sensor-cell array of the camera describes (in some abstract sense) a cylindric surface. The *image cylinder* describes the

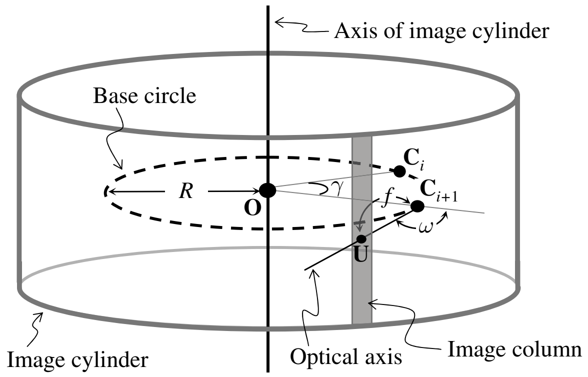

Figure 5: Basic entities of a rotating sensor-line camera.

mathematic abstract location of those rotating tri-linear sensor-lines. Parameters M and L are used to describe the *size of a panoramic image*, captured by a rotating sensor-line camera, where M denotes the number of pixel sensors in the line, and L denotes the total number of lines captured for generating this panoramic image.

The rotation axis is the axis of the image cylinder, and point $\mathbf{O}$ on the axis denotes the center of base circle. The base circle has a radius R, which is called the *off-axis distance*. The optical axis of a camera at position $\mathbf{C}_i$ forms a *principle angle* ω with the ray emitting from $\mathbf{O}$ and passing through $\mathbf{C}_i$ (see Figure 5). The angle defined by two adjacent camera positions, i.e., $\angle \mathbf{C}_i\mathbf{O}\mathbf{C}_{i+1}$, is called *angular increment* and denoted by γ. Moreover, $\mathbf{U}$ defines the point where the optical axis intersects with the image cylinder. The Euclidean distance between $\mathbf{C}_i$ and $\mathbf{U}$ identifies the *focal length* f of the camera at position $\mathbf{C}_i$. In the ideal case, the focal length f, the principle angle ω, and the angular increment γ are assumed to remain constant during a rotation of a sensor-line camera (i.e., during the recording of one panoramic image).

This model generalizes various panoramic imaging models [Ishiguro et al. 1992, Murray 1995, Li et al. 2004]. The four intrinsic sensor parameters, R, f, ω, and L characterize how a panoramic image is acquired. Consider two panoramas, $E_{\mathcal{P}_1}$ and $E_{\mathcal{P}_2}$. The geometric relationship between both sensor coordinate systems can be described by a 3×3 rotation matrix $\mathbf{R}$ and a 3×1 translation vector $\mathbf{T}$. The rotation matrix is given by three row vectors $[\mathbf{r}_1^T \mathbf{r}_2^T \mathbf{r}_3^T]^T$, and the translation vector equals $(t_x, t_y, t_z)^T$.

Of course, when dealing with real rotating sensor-line cameras in real-world applications, we have to understand deviations from those assumptions. Nevertheless, we understand that the sensor (pre-)calibration is very important for

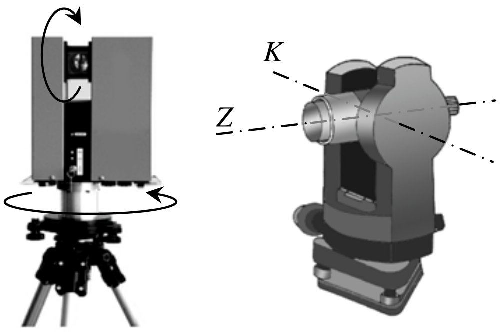

Figure 6: Sketch of scan geometry defined by two scan directions (vertical and horizontal): laser range-scanner (left), and a theodolite (right) with two rotation axes, traditionally called K (German: Kippachse) and Z (German: Zielachse) for such a device.

this approach, and we explain our method in the next section.

Laser scanners differ with respect to their scan geometry (i.e., how the scan rays are progressing during a single scan of a 3D scene). Our range-scans were typically recorded by a panoramic scanner Z+F IMAGER 5003. In this case, the scan geometry is defined as follows: angular increments between scan rays are uniformly defined in two dimensions, which are vertically by a rotating deflecting mirror, and horizontally by rotating the whole measuring system. The vertical

Figure 7: Data of a calibrated LRF image in spherical coordinates.

scan range of the IMAGER 5003 is 310° (which leaves 50° uncovered), and the horizontal scan range is 360°. This scan geometry is similar to the one known for theodolites (see right of Figure 6), which are traditional instruments for measuring (manually) both horizontal and vertical angles.

Figure 7 shows a raw data set without redundancy (i.e., 310° times 180°) captured by using the described panoramic scan geometry. This is the degenerate case of our sensor model that it has $R = 0$, and hence ω becomes meaningless in this case.

2.2 Coordinate Systems for Sensors

In case of a camera with a single sensor-line, we use index j to identify different pixel locations. An *image vector* $\mathbf{v}_j$ points from the (current) camera's projection center $\mathbf{C}$ to the image point (sensor element, pixel) of index j. We have

$$\mathbf{v}_j = \begin{pmatrix} 0 \\ j\tau - y_0 \\ f \end{pmatrix}$$

where τ is the height of the pixel (assuming squared pixel), and y_0 denotes the image center (intersection point of sensor-line with the camera's optical axis).

A local 3D sensor coordinate system (with origin at $\mathbf{O}$) is used to describe the orientation and position of the sensor system in relation to a defined world coordinate system (with origin at $\mathbf{W}$). The Y_o-axis of the sensor coordinate system coincides with the rotation axis (pointing downward; see Figure 8). Let $\mathbf{R}$ denote the rotation matrix, and $\mathbf{t}_0$ denotes the translation vector between sensor and world coordinates systems.

Rotation angle φ is defined to be the angle between the Z_o-axis and line segment $\overline{\mathbf{OC}}$. A rotation matrix $\mathbf{R}_{\varphi(i)}$ is used to describe the camera's orientation at $\mathbf{C}_i$ with respect to the local sensor coordinate system.

A 3D point with respect to the world coordinate system $\mathbf{P}_w$ can be expressed by its corresponding image vector $\mathbf{v}_j$ as follows:

$$\mathbf{P}_w = \mathbf{t}_0 + \mathbf{R}\mathbf{R}_{\varphi(i)} \left[\lambda \mathbf{R}_\omega \begin{pmatrix} 0 \\ j\tau - y_0 \\ f \end{pmatrix} + R \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right]$$

where matrix $\mathbf{R}_\omega$ specifies the additional rotation of the sensor-line when $\omega \neq 0$.

Figure 9 illustrates the general case of a rotating red-green-blue sensor-line camera. In applications we also have to model the following deviations:[4]

[4] Optic distortions are dealt-with for a given camera in a pre-calibration process which is not part of the geometric correction process.

- At any discrete moment i of time, the sensor-line is tilted (within the local coordinate system) by three angles which define a time-dependent rotation matrix $\mathbf{R}_i(\alpha, \beta, \delta)$; this defines the *inner pose* of a sensor-line about all three axes with respect to the central point x_0, y_0.
- The red and blue "sub-lines" have an offset Δ with respect to the central point on the green line.
- The optical axis is rotated by ξ about the X_o-axis.
- The optical axis is rotated by the fixed principle angle ω about the Y_o-axis.
- The sensor-line is rotating with (an eccentricity, or a desired) off-axis distance $R > 0$.

The inner pose $\mathbf{R}_i(\alpha, \beta, \delta)$, the central point (x_0, y_0), and the off-set Δ allows now to decide the positioning of the sensor line in any case. The image vector $\mathbf{v}_j$ is split into two terms as follows:

$$\mathbf{v}_j = \mathbf{v}_{j,\Delta} + \mathbf{v}_f = (\Delta_x - x_0, j\tau + \Delta_y - y_0, 0)^T + (0, 0, f)^T$$

Altogether, the coordinate transform is now the following:

$$\mathbf{P}_w = \mathbf{t}_0 + \mathbf{R}\mathbf{R}_{\varphi(i)} \left[\lambda \mathbf{R}_\xi \mathbf{R}_\omega \left[\mathbf{R}_i \begin{pmatrix} \Delta_x - x_0 \\ j\tau + \Delta_y - y_0 \\ \Delta_z \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ f \end{pmatrix} \right] + \begin{pmatrix} 0 \\ 0 \\ R \end{pmatrix} \right] \quad (1)$$

This equation is used for sensor calibration.

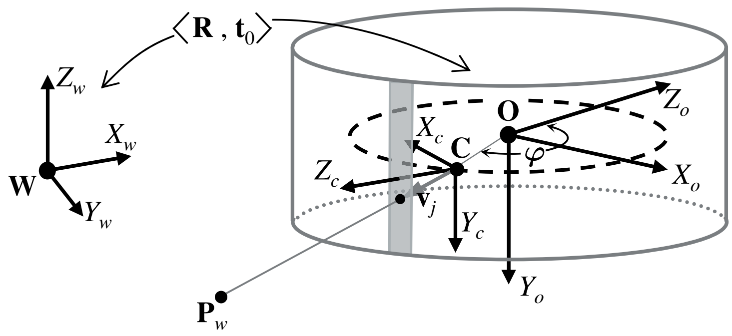

Figure 8: Local sensor coordinate system (with origin at $\mathbf{O}$) and camera coordinate system (with origin at $\mathbf{C}$).

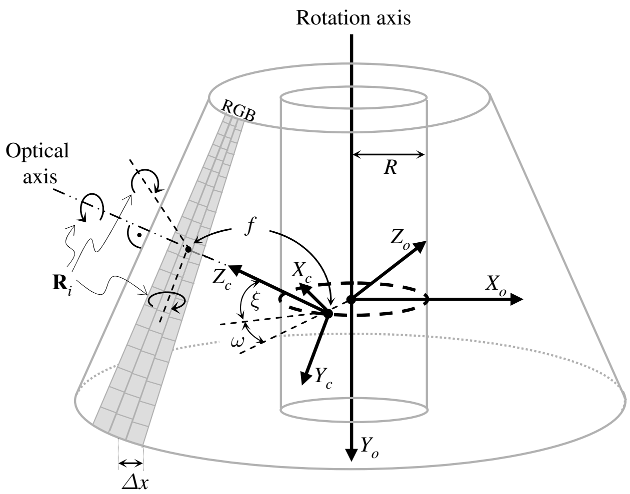

Figure 9: The sensor coordinate system of the rotating line camera: the optical axis identifies the central point x_0, y_0 and is tilted by $\mathbf{R}_\xi$ and $\mathbf{R}_\omega$; each red-green-blue line has a constant distance Δ between central point (on green line) and red or blue line. The tilt of the sensor-line with respect to the optical axis is specified by $\mathbf{R}_i(\alpha, \beta, \delta)$.

3 Sensor Calibration

The common (and straightforward) camera or sensor parameter calibration approach is point-based approach. It is to minimize the difference between ideal and actual projections of known 3D points, such as calibration marks on a calibration object, or localized points in the 3D scene. By taking many images of calibration marks, we are able to apply a least-square error (LSE) optimization procedure.

3.1 Parameters and Objective Functions

In the sequel we describe a least-square approach, as known from photogrammetry, but adapted by us to a panoramic sensor. This approach determines unknown extrinsic parameters of the sensor, which are matrices $\mathbf{R}$, $\mathbf{R}_\xi$, $\mathbf{t}_0$, and off-axis distance R and principle angle ω.

It also determines the intrinsic parameters, which are matrix $\mathbf{R}_i(\alpha, \beta, \delta)$, describing the tilt of the sensor, the “focal length” $f + z_0$, and the sensor's central point x_0, y_0; the latter one also written as vector Δ.

The rotation angle φ of the rotating sensor (sensor-line camera or laser range-finder) may be measured using an internal measuring system of the turntable. Modern technology allows that the angle (for each vertical scan line of the sensor) is determined with an accuracy of 1/1000 degree at least.

Note that the frequently needed recalculation of a "focal length" (i.e., of the camera constant) aims at an exact determination of the (typically unknown) virtual projection center of a pinhole-type model, namely the distance between the entrance pupil to a virtual sensor plane which fulfills the linear imaging assumption.

An *observation* is a recorded calibration mark (with physically measured coordinates, identified with a point (X, Y, Z) such as, e.g., the centroid of the mark) at corresponding image coordinates i and j (i.e., pixel (i, j) for the rotating sensor-line camera, when projecting point (X, Y, Z) into the cylindric panorama). Note that two observations are derivable for one calibration mark because of using two collinearity equations (i.e., one observation is given by two collinearity equation and its corresponding residues).

We have a linear system of n equations with m unknown; the s^{th} observation is given by l_s. The sum of all equations can be written in this form:

$$\sum_{s=0}^{n} l_s = a_{11} \cdot x_1 + a_{12} \cdot x_2 + \ldots + a_{sm} \cdot x_m$$

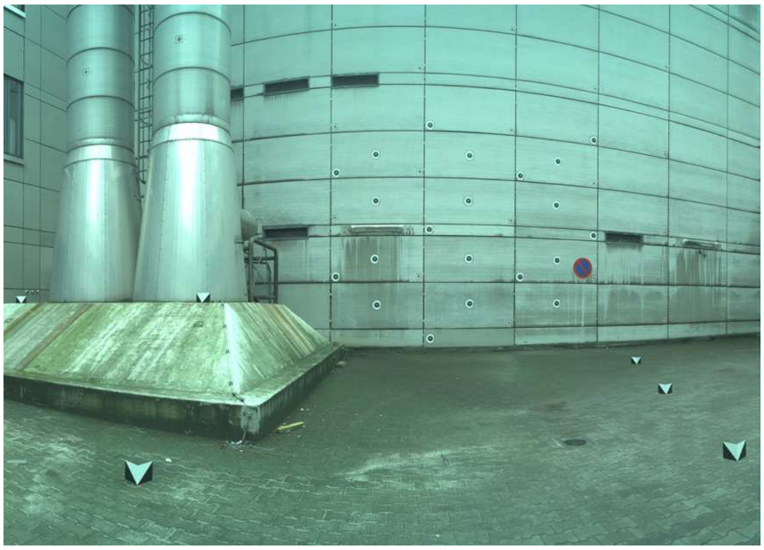

Figure 10: Calibration courtyard at the Institute for Photogrammetry at TFH Berlin. Calibration marks are distributed within the scene.

Observations are considered to be the residues of an iterative Taylor approximation of kth order (which defines a Newton method):

$$l = F(u) - \nabla F^k(\hat{u})\Delta u$$

For the determination of extrinsic parameters and the calibration of intrinsic parameters of a sensor, we place various calibration marks "around the sensor" in the scene (see example in Figure 10). Some of them are projected into image data (depending on visibility), and we assume that all projected calibration marks can be uniquely identified in resulting image data (e.g., in the panoramic image).

Assume that we have m unknowns in total (i.e., elements in matrices, vectors, and parameters), and given are n observations, with $n \geq m$.

3.2 General Error Criterion

We use Equation (1) to model the geometric mapping of 3D points into the sensor coordinate system. By substituting $\mathbf{A} = \mathbf{R}\mathbf{R}_{\varphi(i)}$, $\mathbf{B} = \mathbf{R}_\xi\mathbf{R}_\omega$, and $\mathbf{C} = \mathbf{B}\mathbf{R}_i$ (with matrix elements $\mathbf{A} = a_{11}, ..., a_{33}$, $\mathbf{B} = b_{11}, ..., b_{33}$ and so forth), where $\mathbf{v}_{j,\Delta}$ is the image vector:

$$\mathbf{v}_{j,\Delta} = \begin{pmatrix} \Delta_x - x_0 \\ j\tau + \Delta_y - y_0 \\ \Delta_z \end{pmatrix}$$

This vector is also written in vectorial components as $\mathbf{v}_{j,\Delta} = (\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z)^T$.

After those substitutions, the general mapping equation is now given as follows:

$$\begin{aligned} \mathbf{P}_w &= \mathbf{t}_0 + \mathbf{A}(\lambda\mathbf{B}(\mathbf{R}_i\mathbf{v}_{j,\Delta} + f\mathbf{z}^\circ) + R\mathbf{z}^\circ) \\ \mathbf{A}^{-1}(\mathbf{P}_w - \mathbf{t}_0) - R\mathbf{z}^\circ &= \lambda\mathbf{C}\mathbf{v}_{j,\Delta} + \mathbf{B}f\mathbf{z}^\circ \end{aligned}$$

We rewrite this for all three components of this equation, using $\tilde{\mathbf{P}} = \mathbf{P}_w - \mathbf{t}_0$:

$$\begin{aligned} a_{11}\tilde{\mathbf{P}}_x + a_{21}\tilde{\mathbf{P}}_y + a_{31}\tilde{\mathbf{P}}_z &= \lambda(c_{11}\mathbf{v}_x + c_{12}\mathbf{v}_y + c_{13}\mathbf{v}_z + b_{13}f) \\ a_{12}\tilde{\mathbf{P}}_x + a_{22}\tilde{\mathbf{P}}_y + a_{32}\tilde{\mathbf{P}}_z &= \lambda(c_{23}\mathbf{v}_x + c_{22}\mathbf{v}_y + c_{23}\mathbf{v}_z + b_{23}f) \\ a_{13}\tilde{\mathbf{P}}_x + a_{23}\tilde{\mathbf{P}}_y + a_{33}\tilde{\mathbf{P}}_z - R &= \lambda(c_{31}\mathbf{v}_x + c_{32}\mathbf{v}_y + c_{33}\mathbf{v}_z + b_{33}f) \end{aligned}$$

The matrix of coefficients $a_{11}, ..., a_{33}$ is finally transposed because of the inversion of the matrix $\mathbf{A}$. (For a rotation matrix we have that $\mathbf{E} = \mathbf{R} \cdot \mathbf{R}^T$ is the unit matrix, and, consequently, $\mathbf{R}^{-1} = \mathbf{R}^T$.)

By dividing these equations we may eliminate the scaling factor λ, and we obtain, from the left-hand sides of those three equations, the following two equations:

$$F_{x/z} := \frac{a_{11}(\tilde{\mathbf{P}}_x) + a_{21}(\tilde{\mathbf{P}}_y) + a_{31}(\tilde{\mathbf{P}}_z)}{a_{13}(\tilde{\mathbf{P}}_x) + a_{23}(\tilde{\mathbf{P}}_y) + a_{33}(\tilde{\mathbf{P}}_z) - R}$$

and

$$F_{y/z} := \frac{a_{12}(\tilde{\mathbf{P}}_x) + a_{22}(\tilde{\mathbf{P}}_y) + a_{32}(\tilde{\mathbf{P}}_z)}{a_{13}(\tilde{\mathbf{P}}_x) + a_{23}(\tilde{\mathbf{P}}_y) + a_{33}(\tilde{\mathbf{P}}_z) - R}$$

For the right-hand sides we obtain that

$$G_{x/z} := \frac{c_{11}\mathbf{v}_x + c_{12}\mathbf{v}_y + c_{13}\mathbf{v}_z + b_{13}f}{c_{31}\mathbf{v}_x + c_{32}\mathbf{v}_y + c_{33}\mathbf{v}_z + b_{33}f}$$

and

$$G_{y/z} := \frac{c_{23}\mathbf{v}_x + c_{22}\mathbf{v}_y + c_{23}\mathbf{v}_z + b_{23}f}{c_{31}\mathbf{v}_x + c_{32}\mathbf{v}_y + c_{33}\mathbf{v}_z + b_{33}f}$$

These are the general collinearities, and we also express them by $F_{x/z} = G_{x/z}$ and $F_{y/z} = G_{y/z}$ in short form.

By linearization of these equations it is now possible to estimate iteratively the unknown parameters

$$\mathbf{u} = (t_{x0}, t_{y0}, t_{z0}, \psi, \phi, \kappa, R)$$

for the left-hand sides $F_{x/z}$ and $F_{y/z}$, and

$$\mathbf{u} = (\xi, \alpha, \beta, \delta, \omega, f, y_0, x_0)$$

for the right-hand sides $G_{x/z}$ and $G_{y/z}$, respectively. The three unknown angle ψ, ϕ, κ specify the rotation angles about the x-, y-, and z-axis, respectively. (Note that $\mathbf{R} = \mathbf{R}_x(\psi)\cdot\mathbf{R}_y(\phi)\cdot\mathbf{R}_z(\kappa)$.) The upper index k is the number of the iteration step. The linearization is given as follows:

$$\nabla(G_{x/z} - F_{x/z}) = \left(\frac{\partial G_{x/z}}{\partial u_1} - \frac{\partial F_{x/z}}{\partial u_1}, \frac{\partial G_{x/z}}{\partial u_2} - \frac{\partial F_{x/z}}{\partial u_2}, ..., \frac{\partial G_{x/z}}{\partial u_m} - \frac{\partial F_{x/z}}{\partial u_m}\right)$$

$$\begin{aligned} F_{x,z}^k - G_{x,z}^k &= \nabla(G_{x,z} - F_{x,z})^k \cdot \Delta\mathbf{u} \\ l &= \mathbf{M} \cdot \Delta\mathbf{u} \end{aligned}$$

For $n = m$, the solution is uniquely given by

$$\Delta\mathbf{u} = \mathbf{M}^{-1} \cdot l$$

assuming linear independence between equations.

For $n > m$ observations (i.e., a typical adjustment problem), we apply now the method of least-square error minimization. The error is given as follows:

$$v = \mathbf{M} \cdot \Delta\hat{\mathbf{u}} - l$$

The error function (which needs to be minimized) is defined as follows:

$$\begin{aligned}\min &= v^T v \\ &= (\mathbf{M} \cdot \Delta\hat{\mathbf{u}} - l)^T (\mathbf{M} \cdot \Delta\hat{\mathbf{u}} - l) \\ &= \Delta\hat{\mathbf{u}}^T \mathbf{M}^T \mathbf{M} \cdot \Delta\hat{\mathbf{u}} - 2l^T \mathbf{M} \cdot \Delta\hat{\mathbf{u}} + l^T l\end{aligned}$$

For identifying the minimum, we differentiate and have the resulting function equal to zero:

$$\frac{\partial \left(v^T v\right)}{\partial \Delta\hat{\mathbf{u}}} = 2\Delta\hat{\mathbf{u}}^T \mathbf{M}^T \mathbf{M} - 2l^T \mathbf{M} = 0$$

This leads to the following solution:

$$\Delta\hat{\mathbf{u}} = \left(\mathbf{M}^T \mathbf{M}\right)^{-1} \mathbf{M}^T l$$

The Jacobian matrix M contains all first-order partial derivatives, and l are the residues as defined above. This is solved by means of iterations; the vector $\Delta\mathbf{u}$ contains the corrections of of each unknown. A minimum is found if the unknowns do not change significantly anymore (e.g., $\sum_{s=0}^{m} |\Delta\mathbf{u}_s| < \varepsilon$, with $\varepsilon = 10^{-9}$).

3.3 Calibration Experiments

Some kind of human intervention is in general required for this calibration approach for identifying the projections of those 3D points in a real scene (e.g., the projected points) used as calibration marks, possibly supported by some SIFT feature detector or moment-based sub-pixel accuracy point locator. If a specially designed calibration object is used, this process can be supported by an automatic calibration mark detection algorithm, where marks are located with sub-pixel accuracy (using, e.g., centroid calculation within a mark's region, or intersection points of approximated straight lines when using a checkerboard).

The described least-square approach was used in many applications of panoramic sensors, and is so far our recommended way for calibrating all the mentioned parameters, possibly also including a tilt of the rotation axis of the sensor.

We report about calibration experiments (for performance evaluation) at the calibration courtyard at the Institute for Photogrammetry of the University of Applied Sciences Berlin (known as TFH Berlin); see Figure 10. The locations of the control points along the buildings can be measured with a theodolite and are photogrammetrically balanced. The deviation of our calculated points to the given control points is shown in Table 1. Further experiments confirmed that the estimation of the parameters of the intrinsic pose is stable what means that it does not change significantly by changing the parameters of the extrinsic pose, and therefore these parameters are properly separated from each other.

Control point	Error in pixel coordinate i	Error in pixel coordinate j
7	0.38	0.03
9	0.09	-0.44
10	0.09	0.08
11	-0.23	0.54
12	-0.17	0.28
14	-0.16	0.05
15	-0.14	-0.37
16	-0.27	-0.35
17	0.26	-0.01
19	0.22	0.30
50	0.01	-0.12
51	-0.09	0.11

Table 1: Example of a typical panoramic adjustment, here for calibration marks of the calibration courtyard at TFH Berlin: The table lists deviations between calculated image coordinates and their actual reference coordinates. All listed values are in subpixel scale.

4 Sensor Pose Estimation

When using panoramic sensors, it is actually standard to aim for a set of leveled panoramas, which means all associated rotation axes have to be parallel (say, all perpendicular to the sea level). It can be achieved by levelers. Figure 11 sketches a leveled pair of sensor.

Panoramas captured by leveled sensors are already, for several years, common tools for virtual navigation [Kang and Desikan 1998] or reconstruction of large scale environments [Ishiguro et al. 1992, Kang and Szeliski 1997]. Leveled panoramas allow large "overlapping" fields of views. The larger the common field

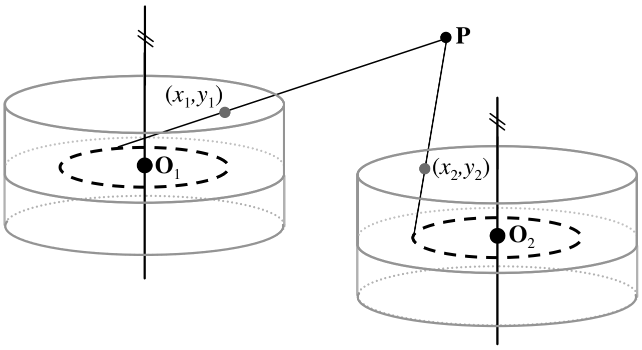

Figure 11: A pair of leveled panoramas and a pair of corresponding image points.

of view, the higher the probability that object surfaces are visible in more than one panorama. Hence, this supports more reliable 3D reconstruction and smooth view-transitions between multiple panoramas in a walk-through simulation.

Sensor pose estimation based on multiple-projection model is a new challenge due to the complexity of their epipolar geometry (i.e., the complex equation of general epipolar curves can be found in [Huang et al. 2008]). The well known linear approaches for camera pose estimation problem such as methods described in [Hartley and Zisserman 2004] are not applicable to our sensor model. The newly proposed approach here for the leveled panoramas is base on the idea of minimizing the distances between the actual image points and their corresponding epipolar curves.

Consider a pair of leveled panoramas acquired by two sensors with the same parameters, and the sensors' poses are related by a single rotation angle ϕ with respect to the rotation axis and a translation vector $(t_x, t_y, t_z)^T$. Given a set of corresponding points (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}), where $i = 1, 2, \ldots, n$, the values of ϕ, t_x, t_y, and t_z can be estimated by minimizing the following sum,

$$\sum_{i=1}^{n}(c_{1i}X_1+c_{2i}X_2+c_{3i}X_3+c_{4i}X_4+c_{5i}X_1X_3+c_{6i}X_1X_4$$
$$+c_{7i}X_1X_5+c_{6i}X_2X_3-c_{5i}X_2X_4+c_{8i}X_2X_5+c_{9i})^2$$

subject to the constraints $X_1^2 + X_2^2 = 1$, $X_1^2 \leq 1$, and $X_2^2 \leq 1$. The five variables to be recovered are $X_1 = \cos\phi$, $X_2 = \sin\phi$, $X_3 = t_x$, $X_4 = t_z$, and $X_5 = t_y$, and the nine coefficients are as follows:

$$\begin{aligned}
c_{1i} &= y_{2i}R\sin(\delta_{1i} - \alpha_{2i}) + y_{1i}R\sin(\delta_{2i} - \alpha_{1i})\\
c_{2i} &= y_{1i}R\cos(\delta_{2i} - \alpha_{1i}) - y_{2i}R\cos(\delta_{1i} - \alpha_{2i})\\
c_{3i} &= -y_{2i}\cos\delta_{1i}\\
c_{4i} &= y_{2i}\sin\delta_{1i}\\
c_{5i} &= y_{1i}\cos\delta_{2i}\\
c_{6i} &= -y_{1i}\sin\delta_{2i}\\
c_{7i} &= f\sin(\alpha_{2i} - \alpha_{1i})\\
c_{8i} &= f\cos(\alpha_{2i} - \alpha_{1i})\\
c_{9i} &= -(y_{1i} + y_{2i})R\sin\omega
\end{aligned}$$

where $\alpha_{ki} = \frac{2\pi x_{ki}}{L}$, $\delta_{ki} = (\alpha_{ki} + \omega)$, and $k = 1$ or 2.

Let (x_1, y_1) and (x_2, y_2) be a pair of corresponding image points in a pair of leveled polycentric panorama $E_{\mathcal{P}_1}$ and $E_{\mathcal{P}_2}$, respectively. Given x_1 and y_1, the

Figure 12: Two symmetric leveled panorama pairs acquired at different locations (top: right panorama of the first pair, bottom: right panorama of the second pair), all marked with 40 corresponding points.

corresponding epipolar curve in $E_{\mathcal{P}_2}$ can be expressed as follows:

$$\begin{aligned}
&y_2 R \sin(\alpha_1 + \omega - \alpha_2 - \phi) - y_2 R \sin\omega \\
&-y_2 \cos(\alpha_1 + \omega) t_x + y_2 \sin(\alpha_1 + \omega) t_z \\
&+f \sin(\alpha_2 - \alpha_1 + \phi) t_y - y_1 R \sin\omega \\
&+y_1 R \sin(\alpha_2 - \alpha_1 + \omega + \phi) + y_1 \cos(\alpha_2 + \omega + \phi) t_x \\
&-y_1 \sin(\alpha_2 + \omega + \phi) t_z = 0,
\end{aligned} \tag{2}$$

The cost function is defined as the row difference between an actual image corresponding point and the point on the same column passing by the epipolar curve. (By some algebraic rearrangements of Equation (2), we my obtain a second-order algebraic representation of these epipolar curves.).

5 Sensor Pose Experiments

Several real-world experiments on estimating sensor poses have been carried out at different places and by using different types of cameras. Camera and the complete sensor were calibrated separately in advance; thus the camera's intrinsic parameters were known and kept unaltered during image acquisition. Figure 12 illustrate one example of a leveled pair taken by a rotating sensor-line camera at different locations in the same room. In this particular example, we used $R = 100$ mm, $f = 21.7$ mm, and $\omega = \pm 155°$. Each panorama has an image

Figure 13: Illustration of three epipolar curves calculated based on the pose estimation results.

resolution of $324 \times 1,343$. A total of 40 corresponding points (marked as stars) were identified for experiments.

The true rotation matrix $\mathbf{R}$ and translation vector $\mathbf{T}$ of these panorama pairs were calibrated with less than $\pm 1\%$ error, and we have that $\phi = 50^\circ$ and $(t_x, t_y, t_z) = (-1,000, -45, -1,000)$ in mm. The estimated sensor pose is denoted as $\hat{\mathbf{R}}$ and $\hat{\mathbf{T}}$. The error measurement for rotation was defined as

$$\arccos\left(\left(tr(\mathbf{R}\hat{\mathbf{R}}^T) - 1\right)/2\right)$$

and the error measurement for translation is defined as

$$\arccos\left(\mathbf{T} \cdot \hat{\mathbf{T}} / \|\mathbf{T}\| \|\hat{\mathbf{T}}\|\right)$$

(i.e., the angle between $\mathbf{T}$ and $\hat{\mathbf{T}}$, both in degrees).

Due to the nonlinear constraints, the quadratic programming optimization approach was not directly applicable. Thus, the sequential quadratic programming method was used instead for optimization (i.e., function *fmincon* in MatLab). In the described example, we obtained 1.22° error in the rotation estimation and 4.65° error in the translation. We show in Fig. 13 three particular epipolar curves calculated based on the erroneous estimations from the leveled case. The average y-difference between the identified corresponding points and the calculated epipolar curves is 1.2 pixel. For most points, the pose estimation errors in this example cause less than three pixel error in vertical direction while processing stereo matching.

We also conducted an error sensitivity analysis with simulated image data, in analogy to the real-world experiment, for both estimation approaches. Figure 14 plots how errors in detecting corresponding points impact the estimation result. The horizontal axis shows various error sizes up to ten pixels. In the analysis, for example, a five-pixel input error means that each pair of corresponding image points was corrupted by errors of max/min five pixels in both x- and y-values, and the errors are modeled as Gaussian-distributed random numbers.

The errors for $\hat{\mathbf{T}}$ are about five times the errors for $\hat{\mathbf{R}}$. From the experiments, we noticed that the assignment of initial values has significant impact onto the estimation result. The estimation result was mostly sensitive to the "sign" of the initial values but not to their quantities nor inter-ratios. In particular, zeros were not good for an initial guess in our case.

Error analysis on R and ω was carried out as well. We concluded that the error of R has a very minor impact on the pose estimation results. Moreover, a k-degree error of ω would cause about a k-degree error in the estimated $\hat{\mathbf{T}}$, for any real number k, but an error in ω has very little impact on the estimation of $\mathbf{R}$.

Finally, more synthetic experiments were designed and performed for different panorama configurations (i.e., different poses, different sensor parameter values, and etc.). They lead to conclusions that the resolution of the input panoramic images, and the distribution of the selected corresponding points are also two critical factors for pose estimation. The panoramic image resolution, especially the width, should be as large as possible. The corresponding points should be distributed uniformly and sparsely on the entire panoramic images. A larger set of corresponding points, say greater than 100, would not guarantee a better estimation result. A much better result can be achieved if image resolution of $1,000 \times 10,000$ is used instead, and the nearest scene point is no less than four meters from both sensors. The estimation errors can be less than 0.5 degrees for both $\mathbf{R}$ and $\mathbf{T}$, allowing for both cases even up to ten-pixel input error.

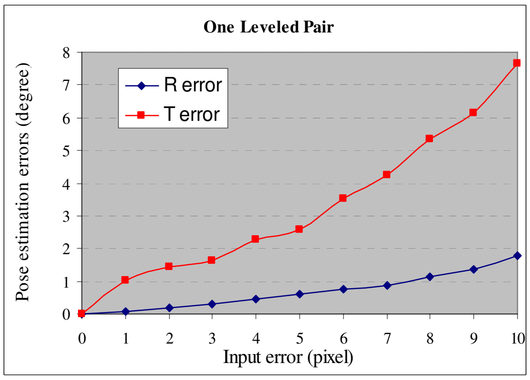

Figure 14: Error sensitivity analysis for the symmetric or leveled case (synthetic images).

6 Conclusions

The paper described the use of two sensor systems (LRF and rotating sensor-line camera) for the generation of accurate and high-resolution 3D models from their originals in the 3D world. Approaches for sensor calibrations and pose estimation were presented followed by some real experiments. Both technologies are essential to ensure accuracy of 3D reconstruction results. We have shown that the proposed approaches are able to achieve high accuracy.

According to our error sensitivity analysis, the sensor calibration results of R and f have very little impact on pose estimation results, while ω's error has a more serious influence on the accuracy of estimated sensors poses. For future work it is thus of interest to develop an algorithm, or a framework, that takes care of sensor calibration and pose estimation at once, similar to self-calibration for the planar image case.

References

[Chen 1995] Chen S. E.: "QuickTimeVR - an image-based approach to virtual environment navigation". In *Proc. SIGGRAPH'95*, pages 29–38, Los Angeles, California, USA, August 1995.

[3D Reality Maps^{TM}] http://www.realitymaps.de/, 2009.

[Boehler et al. 2003] Boehler, W., Vincent, M.B., Marbs, A.: "Investigating laser scanner accuracy". In Proc. Int. Symp. CIPA, pages 696–701, 2003.

[Daniilides and Klette 2006] Daniilidis, K., Klette, R. (editors): "Computer Vision Beyond the Pinhole Camera"; Springer, Amsterdam, 2006.

[EISATS] *.enpeda..* Image Sequence Analysis Test Site (EISATS): `http://www.mi.auckland.ac.nz/EISATS`, 2009.

[Hartley and Zisserman 2004] Hartley, R.I., Zisserman, A.: "Multiple View Geometry in Computer Vision"; Cambridge University Press, second edition, 2004.

[Huang et al. 2008] Huang, F., Klette, R., Scheibe, K.: "Panoramic Imaging: Sensor-Line Cameras and Laser Range-Finders"; Wiley, Chichester, U.K., 2008.

[Ishiguro et al. 1992] Ishiguro, H., Yamamoto, M., Tsuji, S.: "Omni-directional Stereo"; *PAMI*, 14(2):257–262, 1992.

[Jiang and Lu 2007] Jiang W., Lu J.: "Panoramic 3D Reconstruction by Fusing Color Intensity and Laser Range Data". IEEJ Transactions on Electronics Information and Systems, 127(4):568–576, 2007.

[Kang and Desikan 1998] Kang, S.B., Desikan, P.: "Virtual navigation of complex scenes using clusters of cylindrical panoramic images". In Graphics Interface, pages 223–232, 1998.

[Kang and Szeliski 1997] Kang, S.B., Szeliski, R.: "3-d scene data recovery using omnidirectional multibaseline stereo". IJCV, 25:167–183, 1997.

[Klette et al. 1998] Klette, R., Schlüns, K., Koschan, A.: "Computer Vision - Three-dimensional Data from Images"; Springer, Singapore, 1998.

[Li et al. 2004] Li, Y., Shum, H.Y., Tang, C.K., Szeliski, R.: "Stereo: reconstruction from multiperspective panoramas". IEEE Transactions on Pattern Analysis and Machine Intelligence, 26:45–62, 2004.

[Murray 1995] Murray, D.: "Recovering range using virtual multicamera stereo". CVIU. 61:285–291, 1995.

[Nayar 1997] Nayar S.: "Catadioptric omnidirectional cameras". In *Proc. CVPR'97*, pages 482–488, San Juan, Puerto Rico, USA, June 1997.

[Parian and Gruen 2004] Parian, A., Gruen, A.: "An advanced sensor model for panoramic cameras." *ISPRS Int. Arch. Photogrammetry Remote Sensing Spatial Information Sciences*, Vol. XXXV, Part B5, pages 24–29, 2004.

[Peleg and Ben-Ezra 1999] Peleg S., Ben-Ezra M.: "Stereo panorama with a single camera". In *Proc. CVPR'99*, pages 395–401, Fort Collins, Colorado, USA, June 1999.

[Reulke and Scheele 1998] Reulke, R., Scheele, M.: "Der Drei-Zeilen CCD-Stereoscanner WAAC: Grundaufbau und Anwendungen in der Photogrammetrie". Photogrammetrie Fernerkundung Geoinformation, 3:157–163, 1998.