# ROTATING LINE CAMERAS: MODEL AND CALIBRATION

By

**Fay Huang**

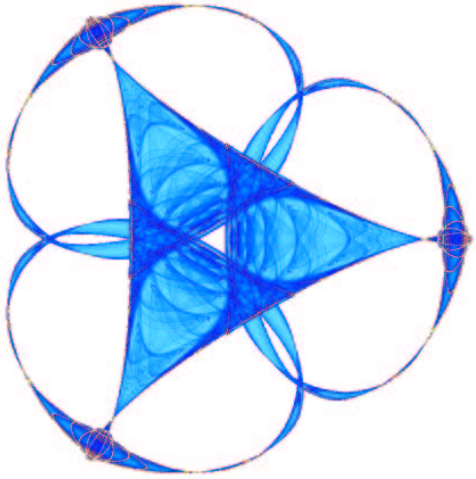
**Shou Kang Wei**

and

**Reinhard Klette**

**IMA Preprint Series # 2104**

( March 2006 )



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall

207 Church Street S.E.

Minneapolis, Minnesota 55455–0436

Phone: 612/624-6066 Fax: 612/626-7370

URL: http://www.ima.umn.edu

# Rotating Line Cameras: Model and Calibration

**Fay Huang, Shou Kang Wei, and Reinhard Klette**

**Taipei and Auckland**

**Abstract.** The studied camera is basically just a line of *pixel sensors*, which can be rotated on a full circle, describing a cylindrical surface this way. During a rotation we take individual shots, line by line. All these line images define a panoramic image on a cylindrical surface. This camera architecture (in contrast to the plane segment of the pinhole camera) comes with new challenges, and this report is about a classification of different models of such cameras and their calibration.

**Acknowledgment.** The authors acknowledge comments, collaboration or support by various students and colleagues at CITR Auckland and DLR Berlin-Adlershof.

# INTRODUCTION

This report is about *panoramas*, which are images of wide viewing angles[1]. Panoramas are either captured with non-pinhole cameras[2] to ensure geometrical correctness of imaged objects for wide viewing angles, or generated from multiple images captured with a pinhole-camera.

For more than a decade, the interest in research on, and developments of panoramic images steadily increased in the computer vision and computer graphics communities (with many popular applications already outside of these communities). Geometric subjects related to panoramic images can be classified into models and the related analysis of panoramic geometry (single-view geometry), or the understanding of geometry of multiple-view panoramic images.

A (single) *view* is a projection of 3D geometry (e.g., an image of a 3D object or a 3D scene) from one viewpoint; it can be monocular or binocular. *Multiple views* not only indicate the existence of multiple single-view images but also a relevance of these images for a potentially improved geometric understanding of a pictured 3D geometry from multiple directions.

## 1.1. Multiple-View Panoramic Images

Multiple views and panoramic images are two options for understanding or visualizing the 3D world. The aim is to combine both for these purposes.

Without any a-priori knowledge (e.g., a model or descriptions of objects in a scene) or additional 3D data (e.g., from a laser range finder or an inertial moment unit), geometric 3D information is rather limited which

[1] For an ideal pinhole camera we have a viewing angle of $\alpha = \arctan(s/2f)$, where $s$ is the size of the photosensitive area in a specified direction (e.g., $24mm$ in vertical or $36mm$ in horizontal direction), and $f$ is the effective focal length, which typically starts at $14mm$, and can go up to multiples of $100mm$. For example, for $s = 36$ and $f = 14$ we have a horizontal viewing angle of about $\alpha_0 = 104.25°$. We define a *wide angle* to be greater than $\alpha_0$.

[2] A *pinhole camera* is defined by central projection onto a plane.

may be inferred from a single-view panorama. As in case of pinhole cameras, it is necessary to have multiple-view panoramas available.

### 1.1.1. MULTIPLE-VIEW GEOMETRY

Over the past decade, there has been a rapid development in the understanding and modeling of the geometry of multiple views in computer vision; see, for example, (Hartley, 1992; Faugeras, 1993; Kanatani, 1993; Luong and Vieville, 1994; Shashua and Werman, 1995; Shashua and Avidan, 1996; Avidan and Shashua, 1997; Heyden, 1997; Hartley and Zisserman, 2000; Faugeras and Luong, 2001; Daniilides and Papanikolopoulos, 2004; Daniilides and Klette, 2006). Most of this work was carried out based on projective geometry and the planar (i.e., the pinhole-camera) image model. Geometric studies of multiple views can be categorized according to the number of views: binocular as in (Hartley, 1992; Faugeras, 1993), trinocular as in (Shashua and Werman, 1995; Avidan and Shashua, 1997), or polynocular (i.e., more than three views) as in (Luong and Vieville, 1994; Shashua and Avidan, 1996).

Epipolar geometry provides the essential fundamentals for all studies about binocular geometry. Epipolar geometry of two (pinhole-camera) views is basically the geometry of the intersections of both image planes with the pencil of planes incident with the *baseline* (i.e., the line joining the optical centers of both cameras). These intersections with the image planes are straight lines, called *epipolar lines*. Consequently, epipolar lines constrain possible locations of corresponding pixels in both images.

The *fundamental matrix* is the algebraic representation of epipolar geometry; see (Laveau and Faugeras, 1994; Zhang et al., 1994; Xu and Zhang, 1996; Zhang and Xu, 1997; Faugeras and Luong, 2001). The fundamental matrix expresses the algebraic relationships between corresponding epipolar lines in both (pinhole-camera) images. The fundamental matrix proved to be useful for *structure from motion*; see, for example, (Pollefeys et al., 1999). Geometric constraints, such as colinearity or coplanarity, for corresponding points can be used to stabilize structure from motion algorithms (Liu et al., 2005).

In generalization of the fundamental matrix, the corresponding algebraic representations for three or four views are called *trifocal* or *quadrifocal tensors*, respectively; see, for example, (Shashua and Werman, 1995; Sawhney and Ayer, 1996; Hartley and Zisserman, 2000; Faugeras and Luong, 2001) for early studies. Besides the geometric information, provided by a trifocal or quadrifocal tensor, the higher complexity of more than two views often becomes an issue with respect to numerical stability or geometric specification of corresponding lines.

Multi-camera heads are a special case of multiple-view pinhole images, see, for example, (Neumann et al., 2002; Valkenburg et al., 2004). Structure from motion or pose recognition benefit from these architectures. See also (Evers-Senne et al., 2004) for the use of a multi-camera-rig, used for reconstructing relatively large 3D scenes.

Core problems in multiple-view geometry include the recovery of the fundamental matrix (or, of the trifocal or quadrifocal tensor) from a set of corresponding image points, the computation of 3D positions of given corresponding points, and camera pose estimation (i.e., determination of relative locations and orientations of cameras from image correspondences).

In conclusion of studies and applications of multiple-view geometry during the past years we can state that more views (images) imply (1) more information (i.e., richer clues about spatial relationships between objects and/or cameras or sensors), (2) higher accuracy (i.e., more visual evidence for geometric inference), and (3) greater visibility (i.e., the total field-of-view increases, and scene occlusions are reduced).

### 1.1.2. PANORAMIC GEOMETRY

A panorama provides a wide field-of-view that may cover different spatial viewing angles, such as $360° \times 360°$ (Nayar and Karmarkar, 2000), a full $360°$ circle (Chen, 1995), or less than $360°$, but still a wide angle. The wide field of view permits that each image contains more information or features so that we have more clues for interrelations between cameras, or between cameras and 3D scenes. Potentially this provides a better support for the inference of 3D world features. The tradeoff is that a larger amount of data is acquired (in comparison to conventional approaches with restrictive fields-of-view). However, as faster computers and larger bandwidth became available, applications and studies using panoramas are now accessible.

The geometry of a single-view panorama is (typically) characterized by a single projection center (the *viewpoint*) and a non-planar image surface (e.g., a cylindric, spheric, or cubic surface). Some approaches also consider multiple projection centers (see Section 1.3 on related work). An apparently advantageous outcome of having two projection centers is that stereovision became possible for panoramas. [Images created at different projection centers are not separate views — as long as they are only temporary results which contribute to one (stereo) panoramic image at one viewpoint. Note that this corresponds to the common understanding of a stereo image for the pinhole camera.]

Multiple-view panoramic geometry contains the case of multiple projection centers in a single-view panorama as a particular case. Epipolar geometry is the intrinsic geometry between two views, or between two projection

centers of a single-view panorama. Epipolar lines in epipolar geometry of pinhole cameras now generalize to parametrized curves (i.e., Jordan curves) in 3D Euclidean space, defined on (cylindric, spheric, quadric and so forth) surfaces of panoramic images.

## 1.2. Motivation and Goal

Panoramic imaging is one of the standard options for visualizing 3D objects or spaces, with a dominating preference for cylindric panoramas. Many applications benefit from using such panoramas; see, for example, (Ishiguro et al., 1992; Chen, 1995; McMillan and Bishop, 1995a; Kang and Desikan, 1997; Kang and Szeliski, 1997; Huang and Hung, 1998; Petty et al., 1998; Wei et al., 1998; Peleg and Ben-Ezra, 1999; Shum and He, 1999; Shum and Szeliski, 1999; Peleg et al., 2000; Huang et al., 2001a; Scheibe et al., 2001; Wei et al., 2002). [We review some of the work in this area (also including applications) in the next section.]

As a result, various types of cylindric panoramic images have been introduced. To ensure sufficient generality, we will not present a geometric study for one particular type of cylindric panoramic images. Our theory will be generic. It is based on a unified representation of panoramic images, which allows to characterize a large range of types of cylindric panoramas. This supports to observe general situations as well as differences within variations of the model.

The study of two-view geometry (note: not binocular geometry!) is a first step towards the more complex geometry of three- or multiple-view panoramas. Epipolar geometry provides (also for these types of images) the fundamentals for understanding two-view panorama geometry. Applications of two-view panorama geometry include 3D scene reconstruction, walk-through animations, stereoscopic visualizations and so forth. The latter two applications are of special interest in todays combined, laser range finder and panoramic image, very-high resolution 3D scene visualizations (Klette and Scheibe, 2005).

A few cases of epipolar geometry for panoramic images have been discussed by (Nene and Nayar, 1998; Svoboda, 1999) — for a pair of mirror-reflection-based panoramas —- by (Gluckman et al., 1998; Shum and Szeliski, 1999) for more specific cases. These studies did not yet contain results for an epipolar geometry analysis for generic panoramic images. The general case was treated in (Huang et al., 2001b).

Previously developed camera calibration methods for the pinhole camera model are not applicable to panoramic images due to the non-linearity of panoramic camera components. Panoramic cameras required a new calibration method, and this was actually one of the harder problems we had

to solve for our experiments. In particular, we were interested in exploring the possibility of using linear features for enhanced numerical stability as well as for computational simplicity.

The reconstruction of 3D scene structure from a pair of stereo images is one of the oldest problems in computer vision, and it is related to work in photogrammetry having its roots in work already done at the end of the 19th century. Spatial sampling of a stereo pair describes how the 3D space is sampled by a stereo camera. This knowledge is essential, because it specifies the potential locations of the 3D samples which can be reconstructed assuming that a perfect image correspondence is established.

### 1.3. Related Work

Applications using panoramic images include areas such as stereoscopic visualization (Huang and Hung, 1998; Peleg and Ben-Ezra, 1999; Wei et al., 1999b), stereo reconstruction (Ishiguro et al., 1992; Murray, 1995; Kang and Szeliski, 1997; Huang, et al. 1999; Shum and Szeliski, 1999; Huang et al., 2001a), walk-through or virtual reality (Chen, 1995; McMillan and Bishop, 1995a; Kang and Desikan, 1997; Szeliski and Shum, 1997; Rademacher and Bishop, 1998; Shum and He, 1999), multimedia and teleconferencing (Rademacher and Bishop, 1998; Nishimura, et al.), localization, route planning or obstacle detection in robot-navigation or mobile vehicle contexts (Yagi and Kawato, 1999; Hong, 1991; Ishiguro et al., 1992; Zheng and Tsuji, 1992; Ollis et al., 1999; Yagi, 1999), tracking and surveillance in 3D space (Ishiguro et al., 1997), and possibly many others. Observable benefits of the use of panoramic images in these applications can be summarized as follows:

**Immersion:** By acquiring a "dense" (with respect to locations of viewpoints) set of panoramas within a scene of interest, one may freely navigate around in a virtual world with a completely 'immersed' sense. Not surprisingly, panorama imaging has also been referred to as *immerse technology*. Note that the 'virtual world' here is different from synthetic scenes generated using computer graphics approaches.

**Interactivity:** Panoramas permit a great degree of interactivity in visualizations, walk-through, multimedia applications, games and so forth, supported by visual consistence with human intuition.

**Simplification:** For example, stereo reconstruction of 3D scenes can be based on complete 360° stereo panoramic images; this bypasses a complicated and erroneous process of merging multiple depth-maps in traditional multiple-pinhole-image approaches.

**Localization:** A robot equipped with panoramic vision can recognize a location by comparing an incoming panorama with a memorized one, and this strategy improves robustness because of more information, provided by a larger visual field.

**Compression:** A panorama composed from an image sequence of a video camera may be seen as a compact version of the original video data, without redundancy.

A number of techniques have been developed for capturing panoramic images of real-world scenes. Based on the characteristics of the imaging systems, we can split these approaches into two categories: scanning-based, or mirror-reflection-based (i.e., *catadioptric*). A scanning-based approach involves a (continuous) motion of a sensor (and possibly also of other parts) of an image acquisition system during the acquisition process. A mirror-reflection-based approach involves one or more mirrors and typically a conventional imaging system (see Figure 1.1); no moving parts are necessary in this case.

We also categorize existing panoramic imaging methods with respect to the number of projection centers associated to a single-view panoramic image. We use two categories, single-center or multiple-center panoramas.

*Figure 1.1.* A commercial catadioptric camera on the 2004 workshop "Computer Vision Beyond the Pinhole Camera".

### 1.3.1. SCANNING-BASED APPROACHES

*Motion-controlled Imaging*
A commonly used acquisition setup for panoramic images is a rotating matrix camera. The camera takes images at constant angular increments (say, of about 20 or 30 degrees); see Figure 1.2(A). A panoramic image is generated by combining (called *stitching*) captured images using steps of registration, merging, and blending (Ayache and Faverjon, 1987; Chen, 1995; Davis, 1998). Error analysis for this kind of panoramic image generation has been characterized and discussed; see, for example, (Kang and Weiss, 1997; Wei et al., 1998a; Chen and Klette, 1999).

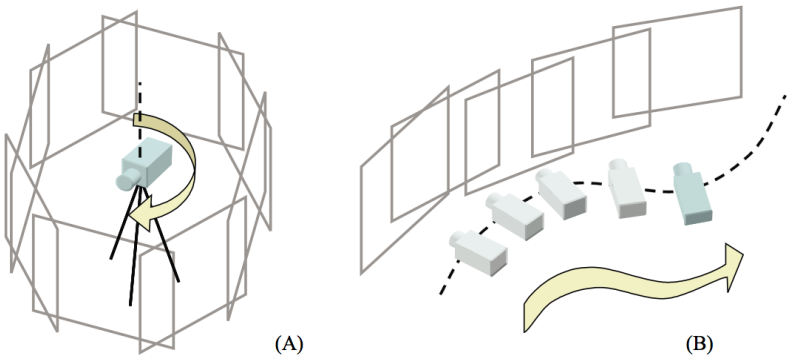



*Figure 1.2.* Scanning approaches: (A) motion-controlled. (B) motion-uncontrolled.

Apple's QuickTimeVR system (Chen, 1995) was one of the first systems to suggest that a traditional CAD-like modeling and rendering process may be skipped. The multi-node version of this system creates environment maps[3] at key locations in a scene. The user is able to navigate discretely from location to location, 'hopping in-between', and to continuously change the viewing direction while at one location (node). This pinhole-camera-based scanning approach has been widely accepted for applications such as virtual travel, real estate, or architectural walk-throughs; see (Irani et al., 1995; Chen, 1995; McMillan and Bishop, 1995a; McMillan and Bishop, 1995; Szeliski, 1996) for its beginnings.

The scanning technique allows that no 3D data is required for rendering views of captured scenes. Additionally, image transformations can be performed uniformly and independently of scene complexities. This technique

[3] Polyhedral environment maps had been proposed by (Greene, 1986) to render either a far away backdrop or reflections of specular objects closer to the viewer.

is today a standard option (even) for real-time or interaction-intensive applications such as 3D games or interactive TV.

(Ishiguro et al., 1992) proposed an image acquisition model that is able to produce panoramas by swiveling a pinhole (video) camera. Details of this approach are discussed in Subsection 1.3.4 for single-view, but multiple-center panoramas. The method is very effective and is popular for multimedia applications. (The model was originally invented for the 3D reconstruction of an indoor environment.) If the resultant panoramic image should be of high quality then the camera motion has to be controlled accurately, and hence the acquisition process is slow. (Ishiguro et al., 1992) already details essential features of a multi-center panoramic image acquisition model, and the approach was further analyzed; see, for example, (Peleg and Ben-Ezra, 1999; Shum and He, 1999; Shum and Szeliski, 1999; Huang et al., 2001c).

*Motion-uncontrolled Imaging*
In computer vision, image mosaicing is often just a single component of a rather complex project (Mann and Picard, 1994; Szeliski, 1994; Anandan et al., 1994; Kumar et al., 1995; Irani et al., 1995; Szeliski and Kang, 1995). As discussed for panoramas above, the extraction of valuable information about a scene typically improves with the collection of more visual data, for example, that of depth (parallax) (Kumar et al., 1994; Sawhney, 1994; Szeliski and Kang, 1995). Unlike the standard panoramic scanning approach mentioned above, an image mosaicing method assumes motion-uncontrolled image sequences (see Fig 1.2(B)), or just sets of images (e.g., taken by different satellites at different days, elevations, and resolutions above the same location). Related disciplines include photogrammetry, computer vision, image processing, and computer graphics, and application areas include change detection, video compression and indexing, or photo editing (Anandan et al., 1994; Irani et al., 1995a; Irani et al., 1995; Lee et al., 1997; Sawhney and Ayer, 1996; Burt and Adelson, 1983). One of the oldest applications is the construction of large aerial views (maps).

(Peleg and Herman, 1997) created panoramic images of 3D scenes from an image sequence of an unconstrained yet continuous camera motion using a mosaicing method called *manifold projection*. Their panoramic image acquisition technique can be characterized by (1) multiple projection centers, (2) flexibility (no calibration, tripod or other tools besides a video camera needed), but also (3) potentially high yet nonuniform resolution. The approach is suitable for 3D scene visualization but not for 3D scene reconstruction.

(Szeliski and Shum, 1997) developed a technique of merging multiple pinhole-camera images, taken at different unknown positions and orienta-

tions. The resultant mosaiced panorama is physically correct if the intrinsic and extrinsic parameters of the camera are completely recovered. However, miss-alignments may still appear in the processed panorama due to various factors [e.g., strong motion-parallax, occlusions, (even) slightly moving objects]. Hence, (Szeliski and Shum, 1997) proposed a deghosting technique following a divide-and-conquer concept (i.e., with respect to image subdivisions and alignments of patches) to minimize miss-registration effects.

(Rademacher and Bishop, 1998) proposed an approach for novel view generation of a 3D object. An object surface point is allowed to appear more than just once in a panoramic image, which is the main difference compared to any other panoramic image technology. This possibly improves the sampling rate of the object (i.e., the same surface point may be sampled for different viewpoints or viewing angles), and may give better quality for the (finally) generated image.

The approach by (Rademacher and Bishop, 1998) requires that the camera motion (i.e., translation and rotation) is continuous in 3D space. The authors also used an inertial moment unit (a motion tracker) and a laser range finder with a rotating mirror. Any miss-registration between the camera motion tracker and the laser range finder produces additional errors in view generation.

### 1.3.2. CATADIOPTRIC IMAGING

*Catadioptric imaging* uses mirrors in combination with a pinhole camera[4] The family of catadioptric panoramas (Zheng and Tsuji, 1992; Gluckman et al., 1998; Baker and Nayar, 1999; Svoboda, 1999) provides real-time and highly portable imaging capabilities at affordable cost. Applications include robot navigation, tele-operation, and 3D scene reconstructions (Daniilides and Klette, 2006). (Yagi et al., 1995) used a conic-shaped mirror for mobile vehicle navigation. (Thomas, 2003) proposed a panoramic camera with a spherical mirror for navigating a mobile vehicle. (Yamazawa et al., 1995) detected obstacles using a panoramic sensor with a hyperbolic mirror. (Zheng and Tsuji, 1992) analyzed features of panoramic images and proposed applications for mobile robot navigation.

Major drawbacks of this approach include low resolution near the image's center, non-uniform spatial sampling, inefficient usage of images (i.e., there is a self-occluded or mirror-occluded area in each captured image), and severe distortions and image blurring due to aberrations caused by coma,

[4] *Dioptrics* is the discipline of refracting elements (lenses), and *Catoptrics* has reflecting surfaces (mirrors) (Hecht and Zajac, 1974) as its subjects. The combination of refracting and reflecting elements is therefore referred to as Catadioptrics.

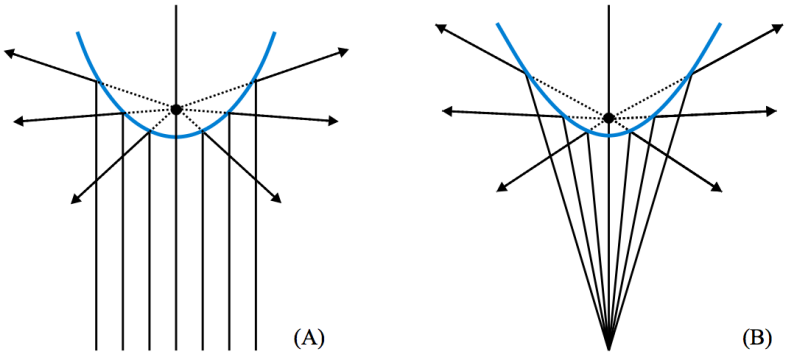



*Figure 1.3.* Catadioptric panoramas: (A) Parabolic mirror with orthographic projection. (B) Hyperbolic mirror with perspective projection (pinhole camera).

astigmatism, field curvature and chromatic aberration. These drawbacks suggest that the catadioptric panoramas are not suitable for applications of recognition or inspection type. They proved to be very efficient in robot navigation, and also for capturing 360 degree panoramas of dynamic scenes (in applications where the drawbacks listed above can be neglected).

(Baker and Nayar, 1999) is an early study of the class of all single-mirror, single-lens imaging systems that satisfy the single viewpoint constraint. There are only two possible combinations which satisfy the single-viewpoint constraint: one is a hyperbolic mirror used in conjunction with a pinhole camera for acquiring 360° panoramas, and the other is the (more theoretic) configuration of a parabolic mirror with an (assumed) orthographic projection camera. Both image acquisition models allow that all the reflected projection rays intersect at a single point (Baker and Nayar, 1999; Svoboda, 1999), and hence possess a simple computational model which supports various applications; see, for example, (Charles, et al.; Nayar, 1998; Yagi and Kawato, 1999; Hong, 1991; Goshtasby and Grover, 1993; Yamazawa et al., 1993; Bogner, 1995; Nalwa, 1996; Nayar, 1997; Chahl and Srinivassan, 1997). The two models are shown in Figure 1.3.

(Svoboda et al., 1998) derived the epipolar curves for both types of single-center catadioptric cameras. The conclusion is that these epipolar curves are general conics for a panoramic pair when using a hyperbolic mirror, and ellipses or lines for a panoramic pair when using a parabolic mirror. (Nene and Nayar, 1998) also studied the epipolar geometry but for the limited case of pure rotation of a hyperbolic mirror, and a pure translation of a parabolic mirror. (Gluckman and Nayar, 1998) estimated ego-motion of catadioptric cameras based on calculated optical flow.

(Nayar and Peri, 1999) introduced a folded catadioptric camera that uses two mirrors and special optics allowing a very compact design. One of the intentions of folded catadioptric cameras is to reduce the size of the entire system. Other works used multiple mirrors and multiple matrix cameras to create more compact imaging systems (Arnspang et al., 1995; Gluckman and Nayar, 1999; Goshtasby and Grover, 1993), or stereo catadioptric panorama imaging systems in the coaxial sense (Nene and Nayar, 1998; Gluckman et al., 1998; Petty et al., 1998; Ollis et al., 1999). Note that if two catadioptric panoramic cameras are specially arranged such that the optical axes are coaxial and each acquired panoramic image is warped onto a cylinder, then the epipolar curves coincide with image columns.

For recent work in catadioptric systems, see (Daniilides and Papanikolopoulos, 2004; Daniilides and Klette, 2006). Catadioptric images are, for example, mapped into cylindric images (with varying resolution along columns: larger on the top and reduced towards the bottom, which corresponds to the center of the mirror). (Zhu et al., 2004) present the collaboration of multiple catadioptric cameras, mounted on mobile platforms, for detecting and tracking human motion; the individual platforms can also calculate their position based on the given panoramic images.

### 1.3.3. SINGLE-CENTER PANORAMAS

Traditionally, a 360° single-center panorama[5] is acquired by rotating a matrix camera with respect to a fixed rotation axis and taking images consecutively at equidistant angles. Around 2000, single-center panoramas taken with respect to multiple-rotation axes (i.e., polycentric panoramas (Huang et al., 2001b)) have emerged and received increasingly interest in applications of 3D scene visualization and reconstruction. For example, (Huang et al., 2001) used a line-scan hyper-resolution camera (Scheibe et al., 2001) (producing about $3 \sim 4$GB data for a single panorama) for acquiring a large number of polycentric panoramas in Auckland or nearby in New Zealand. A (low-resolution) copy of one of those images is shown in Figure 1.4.

Unlike 3D reconstruction with multiple planar images, which usually only cover a narrow field-of-view of a scene, a multiple-view, single-center panorama approach supports omnidirectional reconstruction of 3D scenes. The approach allows bypassing the need for merging of numerous (pinhole camera) depth-maps. However, the approach does not simplify tasks in

[5] The model of a single-center panoramic image combines a single focal point with a cylindric or spheric image surface. All the pixel data in a panoramic image are collected through projections with respect to this single projection center.

*Figure 1.4.* Mahurangi West. A 360° panorama of about 3.5GB, taken with a rotating line camera in 2001. This very beautiful image shows mostly a sunny island scenery, but also (in a high-resolution version of this image) rain going down over one area.

stereo matching, even if camera motion (during the rotation) is well controlled. The epipolar curves are basically sine curves (McMillan and Bishop, 1995a; Kang and Szeliski, 1997; Wei et al., 1999; Wei et al., 1999a; Huang et al., 2001b). This introduces additional uncertainties into stereo matching due to a resampling of pixels along sine curves on discrete images. (Kang and Desikan, 1997) have reported that the matching quality at pixel positions close to the epipoles is very unreliable. They suggested that at least three non-collinear panoramas should be acquired for each 'cluster' in a 3D scene.

Hyperbolic or parabolic catadioptric cameras are the only panoramic catadioptric cameras which have a single projection center. Epipolar curves are general conics for a hyperbolic camera, and ellipses or lines for a parabolic camera. (Wei et al., 2000) attempted stereo matching along such epipolar curves for pairs of catadioptric panoramas. The sampling rate, the shape of the support-region, and the separation between neighboring epipolar curves have been analyzed. The used optimization method was dynamic programming (Gimelfarb et al., 1972; Ohta and Kanade, 1985; Gimelfarb, 2002), and the cost function was defined by correlations of support-regions. According to their results, collecting sufficient and representative samples over epipolar curves is one of the key factors for achieving good quality matching results.

Approaches for stereo matching in catadioptric panoramas typically focus on the coaxial case. In the coaxial case, the epipolar lines are straight and emit radially from the epipole to the boundary of the image. (Gluckman et al., 1998) used a coaxial catadioptric panorama pair and a correlation method with a rectangular support-region. (Ollis et al., 1999) studied various coaxial configurations of catadioptric panoramic image-acquisition models. The results are similar to those of (Gluckman et al., 1998); for example, a skewed rectangular shape of the support-region gives better matching results than a rectangular shape.

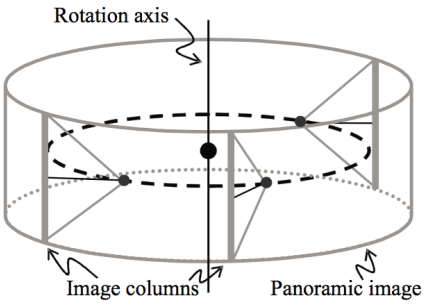



*Figure 1.5.* Example of a multiple-center panoramic camera model.

### 1.3.4. MULTIPLE-CENTER PANORAMAS

A new family of panoramic acquisition models received much attention, starting in the late 1990*s*, in applications of 3D scene visualizations and reconstructions (Peleg and Herman, 1997a; Wei et al., 1998; Peleg and Ben-Ezra, 1999; Rademacher and Bishop, 1998; Shum et al., 1999; Shum and Szeliski, 1999; Shum and He, 1999; Huang et al., 2001a; Huang et al., 2001b; Wei et al., 2002). The new panoramic acquisition model is now characterized by multiple centers (i.e., a single panorama is captured with more than just one projection center). Several disjoint regions (e.g., each defined by several or at least one image column) of a panoramic image are associated with distinct projection centers, one center for each region.

For example, all the $N$ columns of a panoramic image may be associated with $N$ projection centers; see Figure 1.5. We assume an ideal camera architecture where a circle is incident with all the projection centers in 3D Euclidean space (i.e., the distance between the rotation axis of the camera and the projection centers remains constant). Chapter 3 presents a more precise definition and comprehensive information about this camera model.

A multiple-projection-center panoramic imaging model has been initially proposed in (Ishiguro et al., 1992) for 3D-scene reconstruction. H. Ishiguro et al. used a single matrix camera, mounted on a mobile robot, swiveling with a fixed rotation axis with constant angular interval increments, and capturing images through a vertical slit-aperture. Placing all the slit images (each is used to produce one image column) side-by-side in capturing order creates a multiple-center panoramic image. Furthermore, they actually used two vertical slit images in symmetry to the rotation axis for generating a stereo pair of panoramic images. A local map (sketching a view from the top) of a computer room was calculated for a given location of a robot by correspondence analysis of the stereo pair. The authors also

discussed route planning using the established local maps, and a method for merging of the previously obtained local maps to generate a final global map (previously also called "environmental map" or "directional indicator"). Their pioneering work has stimulated the definition of the multiple-center panorama acquisition model as used in our work in the previous years, and also in this report.

(Peleg and Ben-Ezra, 1999) described a model using circular projections (see Chapter 2 for further explanations) for stereo panoramic image generation, which allows a left and a right eye perception of panoramic images. The left and right images are approximated with respect to views obtained from the inner circle of their cylindrical model. (Shum and He, 1999) proposed a concentric model of panoramic images for image-based rendering in which novel views within an inner circle, and between the inner and outer circles are approximated by circular projections in normal direction [the same as in (Peleg and Ben-Ezra, 1999)] and in tangential direction. A more general analysis of the concentric model, from 2D-camera motion to 3D reconstruction, has also been discussed in (Shum et al., 1999).

# CAMERA MODEL

A panorama may cover spatial viewing angles of a sphere [i.e., 360° × 360°, see (Nayar and Karmarkar, 2000)], a full 360° circle (Chen, 1995), less than 360°, but still a wide viewing angle, or even more than 360° if time also defines a dimension. There are various geometric forms of panoramic sensor areas. This report focuses on cylindric 360° panoramas.

The projection center of a single-center panorama camera is known as the *nodal point*, because it is the intersection of optical axis and rotation axis. This chapter presents the definition and main characteristics of a general multi-center panorama camera. We investigate a straightforward realization of panorama acquisition based on this camera model, followed by characterizations of a few selected geometric configurations of multiple-view panoramas, each exemplifying the basic multi-center camera model. Special attention is given to existing applications, practical advantages, and possible generalizations of the model.

## 2.1. Basic Model

The camera model is an abstraction; we exclude modeling of mechanical or optical errors (e.g., we do not consider lens distortions) and restrict ourself on a simple model defining the projective geometry.[1]

A *projection center* is also known as *optical center* or *focal point*, and the *sensor area* of a camera is specified by a light-sensitive 2D manifold (given by the shape of a photon-sensing device, which is a rectangular planar array of CCD cells in case of a pinhole camera). Our camera model has multiple projection centers and a cylindrical sensor area; see Figure 2.1. "Sensor area segments" are captured, one for each projection center. A

[1] One reason for doing so is that the used cameras were accurately calibrated at a special facility at DLR Berlin with respect to camera geometry and photometric properties; there remain critical issues in our abstraction such as the assumption of ideal circular rotation of the sensor; however this proved to be not (yet) of relevance for our studies so far.

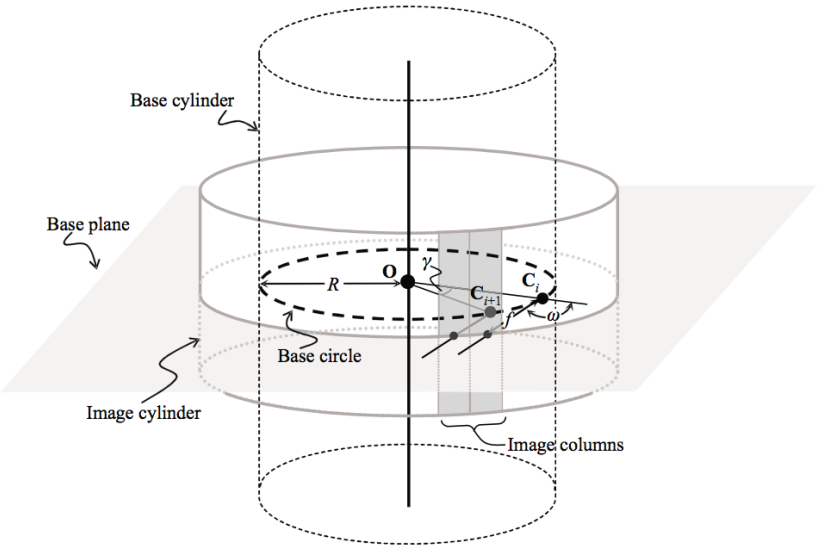



*Figure 2.1.* Camera model and geometric projection. Parameter $f$ is explained in the context of Figure 2.2.

resulting panoramic image is composed of data captured by these sensor area segments (as described below)[2].

**C** (with subscript $i$) denotes in the figure one of the projection centers. Projection centers are uniformly distributed on a circle called *base circle*, which is shown as a bold-dashed curve. The plane which is incident with the base circle, is called *base plane*. **O** denotes the center of the base circle and $R$ is the radius of the base circle. $R$ is called the *off-axis distance*. Image segments are captured on the shown straight cylindric surface of finite height, which is called the *image sensor cylinder*, or short the *image cylinder* (e.g., in the sense of a captured panoramic image). The center of the base circle coincides with the center of the image sensor cylinder. The normal vectors of the base circle (in the base plane) are perpendicular to the axis of the image sensor cylinder. The tall dashed cylinder (in the figure) of radius $R$ (which is is coaxial to the image cylinder) is the *base*

[2] This report will not discuss the issue of different capturing times for these sensor area segments; typically, segments capture data in clockwise or counter-clockwise order, one segment at a time, which defines limitations for the representation of dynamic scenes: despite of having fast cameras, moving objects *are* distorted in resulting panoramic images. Regarding dynamic scenes, created 3D representations of static scenes can, for example, be used as animation backgrounds. However, catadioptric panoramic images are better suited for capturing (!) dynamic scenes.

*cylinder*.

An angle $\gamma$ describes the constant angular distance between any pair of adjacent projection centers on the base circle with respect to $\mathbf{O}$, and it is called the *angular unit* (illustrated by $\mathbf{C}_i$ and $\mathbf{C}_{i+1}$ in Figure 2.1; formally $\gamma = \angle \mathbf{C}_i\mathbf{O}\mathbf{C}_{i+1}$). We have $\gamma = 0$ in the theoretical case that any point on the base circle acts as a projection center. An optimized choice of the value of $\gamma$ according to the distance between the scene of interest (defined by the closest object) and the camera is discussed in (Wei et al., 2002).

As default we assume a finite number of projection centers; the image sensor cylinder is partitioned into parallel stripes of equal width that are parallel to the axis of the image sensor cylinder. These parallel stripes are *sensor columns* (or *image columns* in a captured panoramic image) of the resulting panoramic camera, and define the "sensor area segments" above. Figure 2.1 shows two adjacent sensor columns (gray-shaded). The number of sensor columns coincides with the number of projection centers on the base circle. It is called the *width* of the resulting cylindric panoramic camera and denoted by $W$. Obviously, we have the following:

$$W = \frac{2\pi}{\gamma}$$

Projection centers and sensor columns are indexed successively following the same orientation. This defines (by having the same index) a one-to-one mapping between projection centers and sensor columns. Indices of projection centers and sensor columns may start with $i = 1$ (independently) at any center or column. Figure 2.2 illustrates such a one-to-one mapping.

Geometrically, we assume for this figure that the line segment from $\mathbf{C}_i$ ends at the center point of the intersection of the base plane with the

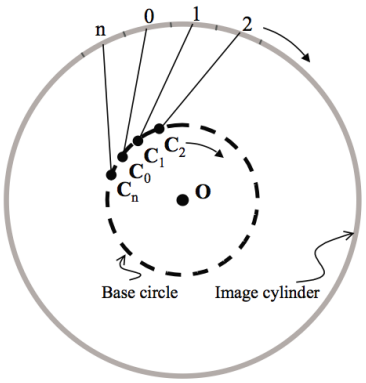



*Figure 2.2.* A mapping between indices of projection centers and sensor columns.

sensor column $i$. All these line segments are of constant length $f$ (see also Figure 2.1). We identify $f$ with the *distance between a projection center and its associated sensor column*. $f$ is the *effective focal length* for this panoramic image.

Besides width $W$ we also define the *height* $H$ of a panoramic image, also called the *number of sensor* or *image rows*. For this we assume uniformly spaced planes, all parallel to the base plane, whose intersection with the image sensor cylinder partitions each sensor column into $H$ patches of equal height. Each patch defines a *sensor element*, also called *pixel* (i.e., the sensor cylinder is composed of $W \times H$ pixels.)

Pixels of cylindric panoramic images are this way either 2D cells (pixels in the *cell model*, as defined above), or each pixel can also be identified[3] with its center point (i.e., centroid) on the image sensor cylinder (pixels in the *point model*). We use the point model as default in the following. Without loss of generality we assume that one of those pixels (called the *center pixel*) is incident with the base plane (for each sensor column). A projection center and its associated sensor column define a bundle of $H$ coplanar projection rays, one for each pixel in this column.

Let $\omega_i$ be the angle formed by a projection ray from projection center $\mathbf{C}_i$ towards the corresponding central pixel of column $i$, and the normal vector of the base circle (at that central pixel; pointing outward). All those angles are constant, $\omega = \omega_i$, for $i = 1, \ldots, W$, and $\omega$ is called the *principal angle* of the panoramic image; see Figure 2.1. Note that $\omega$ is defined starting from the normal of the base circle in clockwise orientation (as seen from the top) over the valid interval $[0, 2\pi)$.

Altogether, $R$, $f$, $\omega$, and $\gamma$ are the defining parameters of a panoramic camera. By choosing these parameters within defined intervals, we define a class of panoramic cameras. (In the real world, such a class would be just one physical camera.) A panoramic image[4] is a value of a function $E_{\mathcal{P}}(R, f, \omega, \gamma, S)$ where parameters specify the used camera and $S$ the captured scene. In general we will not list $S$, just $E_{\mathcal{P}}(R, f, \omega, \gamma)$. – The notations defined in this section are used throughout the report.

## 2.2. Progressive Scanning Approach

The implementation of the defined camera model was a subject in (Peleg et al., 2000). The authors proposed two theoretical solutions: a spiral-mirror, or a spiral-lens design for real-time cylindric panorama imaging. Unfortunately, the spiral-mirror approach cannot cover a complete circular 360°

[3] In analogy to digital geometry (Klette and Rosenfeld, 2004) of planar images.

[4] We follow (Klette et al., 1998) where images have been identified with irradiances, thus using the standard symbol $E$ of irradiance for an image.

*Figure 2.3.* Three people ("ghosts") move when the rotating line sensor was scanning in their direction.

viewing angle. Furthermore, both approaches assume constant $R$ and $\omega$ values so that these methods cannot deal with more variable specifications of panoramic cameras.

This section discusses a progressive scanning approach. Two off-the-shelf (or commercially available) camera architectures are discussed, namely line cameras and matrix cameras, both characterized by specific sensor areas. The basic idea of the progressive scanning approach is to approximate the given camera model geometrically by rotating a line or matrix camera, assuming a fixed rotation axis and capturing image segments consecutively at equidistant angles, angularly spaced by $\gamma$, up to a complete 360° scan. Each

*Figure 2.4.* Top: a 360° image from Auckland's harbor bridge in 2001 (a rotational scan which took about 4 min; the image is about 350 Megapixel). Bottom, left to right: steps of zooming into the image above, which shows finally (on the right) image jitter which is due to the vibration of the bridge during scanning.

captured image segment contributes one image column to the panoramic image. During a complete 360° scan, the path of the optical centers (i.e., projection centers) forms a base circle, and the image sensor surface defined by contributing image columns forms an image sensor cylinder.

This approach allows complete 360°-scans, but cannot be applied for dynamic scenes (see Figure 2.3). Due to changes in viewing angles, photometric properties of captured panoramic images also differ from those of pinhole-camera images. This can be of artistic value, and defines also a new research subject. The image brightness at both ends of a scanned panorama can also differ due to varying illumination (as common for outdoor images). Finally, vibrating tripod (see Figure 2.4) creates image jitter.

### 2.2.1. SINGLE-LINE CAMERA

Single or multiple-line cameras have been designed and produced for push-broom image capturing in aerial photogrammetry, and, as a spin-off, they led to rotational, single-line cameras (Scheibe et al., 2001).

The basic architecture of a single line camera (assuming the same flexibility as in our camera model) is illustrated in Figure 2.5(A). Its geometry would be characterized by an optical (projection) center, a 1D linear photon-sensing device (e.g., CCD), an effective focal length $f$, and (when rotating this configuration) an off-axis distance $R$ and a principal angle $\omega$. (Rotation is normally supported by a 360°-capable turntable, or a rotation rig on a tripod.)

If the optical center lies on the rotation axis of the turntable (i.e., $R = 0$) and the sensor line is parallel to the rotation axis [as for the camera described in (Scheibe et al., 2001)], then the geometry of the resulting panorama is characterized by a single projection center.

If $R > 0$ then multiple projection centers generate a path that approximates a base circle. Before starting a progressive scan, the off-axial single-line camera may rotate about its optical center (while the sensor line

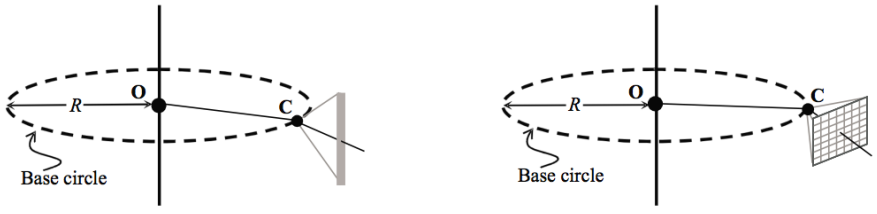



*Figure 2.5.* Camera models: (A) single-line camera and (B) matrix camera.

is kept parallel to the rotation axis) into a fixed angle $\omega$, with $0 \leq \omega \leq 2\pi$. Then a full 360°-scan is performed, where the fixed angle $\omega$ defines a particular view.

### 2.2.2. MATRIX CAMERA

A cylindrical panorama can also be scanned progressively by a pinhole (matrix) camera; see Figure 2.5(B). We proceed as described for the single-line camera except that the captured image line is now taken from a fixed sensor column of the sensor matrix. The choice of the column corresponds (somehow; see below) to the choice of angle $\omega$ in case of the single-line camera.

Similar to the line-camera case, if the optical center of the matrix camera coincides with the rotation center of the turntable, and the sensor columns are parallel to the rotation axis, then the resultant panoramas are single-projection-center panoramas. Ignoring optical distortions, in this case any sensor column can be chosen, and it would always be the same generated panoramic image (assuming a static scene). Moreover, when a matrix camera is placed eccentrically from the rotation axis of the turntable or rotation rig (i.e., $R > 0$), the scanned panoramas are multiple-projection-center panoramas.

However, in difference to the single-line camera case, a matrix camera of image resolution $H(\text{height}) \times W(\text{width})$ is able to generate $W$ different panoramas after a complete 360° scan. Furthermore, all these panoramas have different effective focal lengths and principal angles, depending on the chosen column.

The panorama composed by the $i$th column of the sensor matrix has the following parameter values: $R$ is equal to the distance between the optical center of the matrix camera and the rotation axis of turntable or rotation rig; the values of the effective focal length and the principal angle are calculated as

$$f = \sqrt{f_{\mathcal{M}}^2 + (i - c_{\mathcal{M}})^2 \mu^2}$$

and

$$\omega = \psi + \arctan\left(\frac{(i - c_{\mathcal{M}})\mu}{f_{\mathcal{M}}}\right)$$

respectively, where $f_{\mathcal{M}}$ is the effective focal length of the matrix camera, $c_{\mathcal{M}}$ is the central sensor column[5] of the matrix camera, $\mu$ is the size of a single photon-sensing element (assuming perfectly square pixels), and $\psi$ is

[5] The central sensor column of a matrix camera is that column where the optical axis of the camera passes through. We assume here (for the formula) that this column is centered in the sensor matrix (and thus uniquely defined).

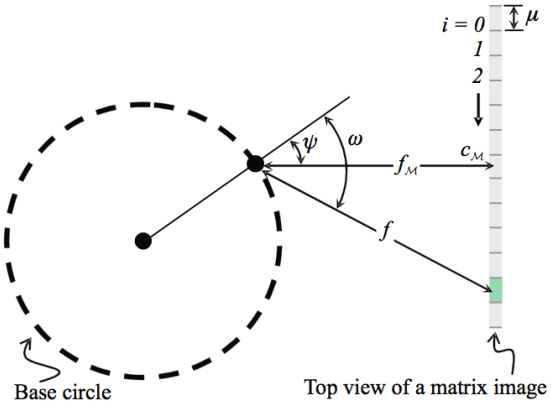



*Figure 2.6.* Calculation of effective focal length $f$ and principal angle $\omega$ when using a pinhole (matrix) camera for panoramic imaging.

the angle between the optical axis of a matrix camera and the normal of the base circle passing through the optical center (see Figure 2.6).

## 2.3. Multiple-view Panoramas

Specific panoramic camera models will exemplify our basic camera model described in Section 2.1. This section describes selected types of multiple-view panoramas (with special attention to existing applications, practical advantages, or possible model generalizations). Types are defined with respect to camera geometries. For these characterizations we use the following convention: each panorama is uniquely associated to an *axis* (i.e., the axis of the image cylinder), and to a *center* (i.e., the center of the base circle).

### 2.3.1. SINGLE-CENTER PANORAMAS

In general, we consider a cylindric panorama as a multiple-projection-center panorama (i.e., $R > 0$); a single-projection-center panorama (i.e., $R = 0$) will only be considered as a special or degenerated case. We assume a finite set of captured panoramas. When all panoramas in this set are captured by a single-center camera, then we call it (a set of) *single-center panoramas*. Today vast applications of 3D scene visualizations are based on, or assume single-center panoramas; see (Zheng and Tsuji, 1992; Chen, 1995; Yagi et al., 1995; Huang et al., 2001) for some early examples.

The main merit a single-center panorama is that it permits the generation of a geometrically correct perspective view on a planar surface (e.g.,

computer screen). The simplicity of the camera model allows fast post-processing and real-time (or interactive) visualization under the assumption of moderate computing power and image resolution. However, it is practically difficult to guarantee that a projection center exactly coincides with the rotation axis. Furthermore, applications under such an over-simplified assumption (i.e., $R = 0$) might be affected by performance instabilities (i.e., image-content dependent performance) or unpredictable error behaviors.

### 2.3.2. POLYCENTRIC PANORAMAS

A set of panoramas with centers at unconstrained positions in 3D space is called (a set of) *polycentric panoramas*. This is our most general case. Camera parameters for each of the polycentric panoramas may differ from one to any of the others. The fully unconstrained case of polycentric panoramas is sketched in Figure 2.7(A).

Polycentric panoramas are defined for studying the geometry of finite sets of cylindric panoramas. Applications will normally define additional constraints, and these are studied as special cases of polycentric panoramas, such as parallel-axes, coaxes, concentric, or symmetric panoramas; see Figure 2.7(B∼E).

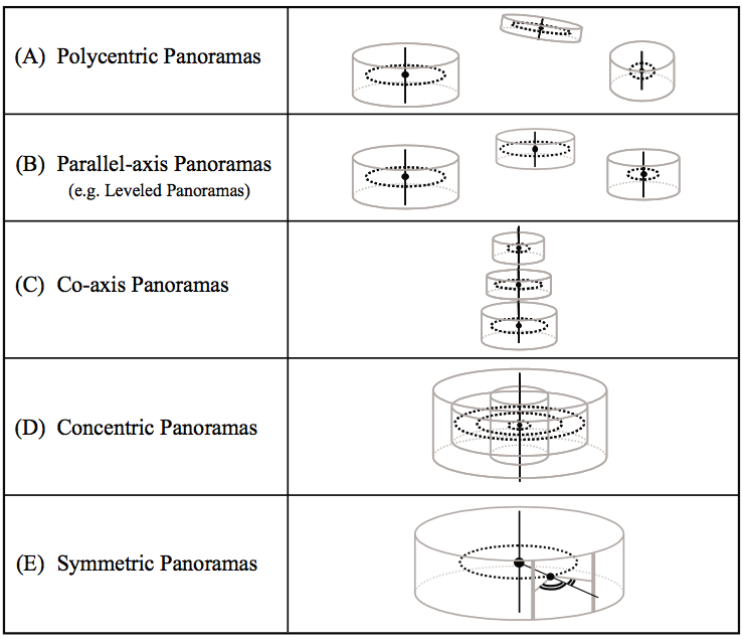



*Figure 2.7.* Different types of finite sets of panoramas.

### 2.3.3. PARALLEL-AXES AND LEVELED PANORAMAS

A set of polycentric panoramas whose associated axes are all parallel is called *parallel-axes panoramas*. Such a set is illustrated in Figure 2.7(B). In particular, if the axes are all orthogonal to the sea level (e.g., by using a round "bull's-eye"), then they are called *leveled panoramas*.

Leveled panoramas are often used for visualization or reconstruction of large areas (e.g., in a museum). There are (at least) four reasons which support their usage:

**1.** Scene objects in resulting panoramas are acquired in natural orientation corresponding to human visual experience.

**2.** The size of overlapping fields of view is maximized in multiple panoramas.

**3.** It is not difficult to ensure them (e.g., by a panoramic camera mounted on a leveled tripod).

**4.** The dimensionality of relative-orientations of multiple panoramas is reduced from three to one dimension.

(The analysis of epipolar geometry will add one more argument.) Obviously, a larger overlapping field of view increases the probability that object surfaces are visible in multiple panoramas, and this supports shape reconstruction or consistent view-transitions among multiple panoramas in a walk-through simulation. The reduction of dimensionality of the relative-orientation problem simplifies panoramic geometry problems, and improves the robustness with respect to inherent errors.

### 2.3.4. COAXES PANORAMAS

A set of polycentric panoramas whose associated axes all coincide is called (a set of) *coaxes panoramas*. Figure 2.7(C) shows an example of a coaxes panorama, composed of three panoramic images which are defined by varying camera parameters and heights of their centers.

If camera parameters of two coaxes panoramas are identical (but different heights of centers), then the epipolar geometry is quite simple. Epipolar lines are corresponding image columns in this case. This simplifies stereo matching. In addition, the implementation of such a configuration is reasonably straightforward (i.e., after positioning a tripod). This configuration is widely used, for example also for catadioptric approaches (Southwell et al., 1996; Nene and Nayar, 1998; Petty et al., 1998).

### 2.3.5. CONCENTRIC PANORAMAS

A set of panoramas where not only their axes coincide but also their associated centers, is called (a set of) *concentric panoramas*; see sketch in Figure 2.7(D). Of course, concentric panoramas can also be leveled (to the sea level), and then all the advantages of leveled panoramas apply as well.

Subsection 2.2.2 discussed the use of a matrix camera for panorama acquisition by progressive scanning. A complete 360° scan of a matrix camera of image resolution $H \times W$ generates $W$ different panoramas, each with individual camera parameters (i.e., a defined effective focal length and a unique principal angle). All $W$ panoramas are concentric.

### 2.3.6. SYMMETRIC PANORAMAS

Two concentric panoramas, $E_{\mathcal{P}_R}(R, f, \omega, \gamma)$ and $E_{\mathcal{P}_L}(R, f, (2\pi - \omega), \gamma)$, are called *symmetric panoramas* or a *symmetric pair*; see Figure 2.7(E). Their principal angles are symmetric to the associated normal vector of the base circle.

A symmetric pair is also called a *standard stereo panorama* or, in short, a *stereo panorama* – in analogy to standard binocular stereo images for pinhole cameras. A feature of stereo panoramas is that the epipolar geometry is characterized by epipolar lines being image rows. Due to this epipolar property, stereo panoramas are directly stereoscopic-viewable. Figure 2.8 illustrates part of an anaglyphic representation of a stereo panorama. [Note: due to the cylindric sensor surface the projection of a horizontal straight line appears curved when displaying the panorama ('incorrectly') on a pla-

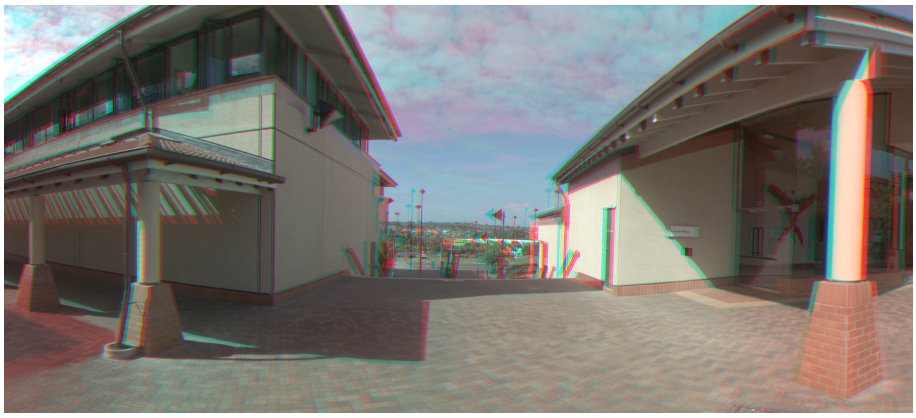


*Figure 2.8.* Part of an anaglyphic panorama (Tamaki Campus, The University of Auckland, 2002). Proper 3D viewing requires anaglyphic eyeglasses.

nar surface.] This property of having simple epipolar lines also supports 3D reconstruction applications; the same stereo-matching algorithms can be used as for standard binocular stereo images (Barnard and Fischler, 1982; Ohta and Kanade, 1985; Cox, 1994; Šára, 1999; Gimelfarb, 2002).

The progressive scanning method described Section 2.2 allows to acquire any of the different types of multiple panoramas. In particular, stereo panoramas can be generated by only one rotational scan using a two-line line camera or a matrix camera. Figure 2.9 shows a prototype stereo panoramic camera based on the wide-angle airborne camera WAAC (Sandau and Eckardt, 1996; Reulke and Scheele, 1998). The camera is fixed on an extended slider, and it is facing towards the turntable. Some experiments reported in the report were conducted using this stereo panoramic camera to acquire various types of multiple panoramas.

Specifications of camera settings or related details are given in the experimental sections of relevant chapters. The complete technical data sheet of WAAC may be found in (Huang et al., 2001). The formulation of the theoretical camera model benefited from these practical experiences, and, vice-versa, we were able to contribute to the design of a new generation of line cameras, which is now widely used.

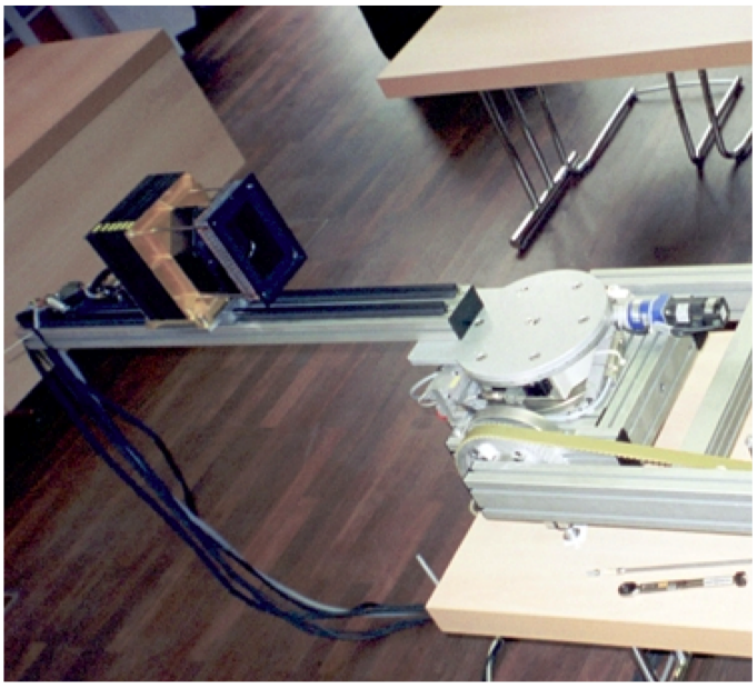


*Figure 2.9.* Stereo panoramic camera at the Institute of Space Sensor Technology and Planetary Exploration, DLR Berlin, 2000.

## 2.4. Coordinate Systems

This section defines the coordinate systems as used throughout the report. The first sections specify those of a panoramic camera; the laser range finder is treated separately in the final subsection.

### 2.4.1. DISCRETE AND EUCLIDEAN IMAGE COORDINATE SYSTEMS

An unfolded (or flattened) panoramic image is of rectangular form of dimensions $W \times H$ (i.e., width $\times$ height) in pixels. It is described in a 2D coordinate system as in digital geometry for the pinhole case (Klette and Rosenfeld, 2004). A pixel has integer coordinates $u$ and $v$ indicating the image column and image row, respectively; see Figure 2.10. Let the origin of the image coordinate system be at the top-left corner of the image. The range of $u$ is from zero to $W$-1, and the range of $v$ is from zero to $H$-1.

We also define a 2D Euclidean image coordinate system for panoramic images. The dimensions of the panoramic image in the continuous space are $W\mu \times H\mu$, where $\mu$ is the physical size (i.e., edge length) of a pixel. [Pixels are assumed to be squares in the cell model.] A pixel (in the point model) has coordinates denoted as $(x, y)$. The $x$-axis is parallel to image rows, and the $y$-axis parallel to image columns; see Figure 2.10. The origin of the image coordinate system is the centroid of the (cell-)pixel $(u, v) = (0, v_c)$, which is also assumed to be incident with the base plane of the panoramic camera.

The relation between discrete and continuous image coordinate system is as follows:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u\mu \\ (v - v_c)\mu \end{bmatrix} \tag{1}$$

where $\mu$ is the pixel size. This chapter assumes that any image is defined on a continuous 2D surface in Euclidean space, having real image coordinates $x$ and $y$. Equation (1) is used when plotting epipolar curves.

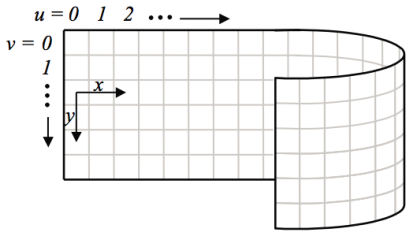



*Figure 2.10.* Unfolded panorama and its discrete and Euclidean image coordinate systems.

### 2.4.2. CAMERA AND OPTICAL COORDINATE SYSTEMS

A panoramic camera defines a 3D Euclidean $X_oY_oZ_o$ coordinate system, called *camera coordinate system*; see Figure 2.11. The origin of a camera coordinate system coincides with the center **O** of the base circle of the panoramic camera. The $Y_o$-axis of a camera coordinate system coincides with the axis of the image cylinder of the panoramic camera. The positive direction of the $Y_o$-axis of a camera coordinate system coincides with the positive direction of the $y$-axis of the image coordinate system. In Figure 2.11, both axes point downward.

The $Z_o$-axis of a camera coordinate system is incident with the projection center associated with the initial image column at $x = 0$. The $X_o$-axis of a camera coordinate system is defined such that $X_oY_oZ_o$ forms a right-hand system. Note that the $X_oZ_o$-plane coincides with the base plane of the camera. The camera coordinate system basically serves as a reference for describing position and orientation of a panoramic camera in a 3D world coordinate system.

We also define a 3D Euclidean $X_cY_cZ_c$ coordinate system for each optical center of the panoramic camera, called *optical coordinate system*. The origin **C** of the optical coordinate system coincides with the projection center of the camera model. $Y_c$-axis of an optical coordinate system and $y$-axis of the camera coordinate system are parallel and point into the same direction; both downwards in Figure 2.11.

The $Z_c$-axis of an optical coordinate system is the *optical axis* for the specified optical center. It is incident with the base plane of the camera and the center of the associated sensor (or image) column. The optical axis of the optical center **C** and the normal of the base circle, incident with **C**,

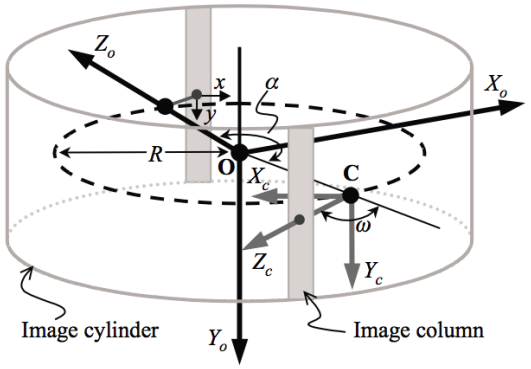



*Figure 2.11.* Camera (bold black lines) or optical (bold gray lines) coordinate systems originate at **O** or **C**, respectively.

form the principal angle $\omega$. The $X_c$-axis of the optical coordinate system is also defined by a right-hand system. The $X_cZ_c$-plane is coplanar to the $X_oZ_o$-plane of the camera coordinate system.

The optical coordinate system is useful for calculating the projection of a 3D point into an image. A 3D point is projected onto an image column iff[6] the X-coordinate of this 3D point with respect to the optical coordinate system is equal to zero.

The relation between an optical coordinate system and the camera coordinate system is as follows:

$$\begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} = \left[\mathbf{R}_{\mathbf{oc}}^{-1}\ \mathbf{T}_{\mathbf{oc}}\right] \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{2}$$

where $\mathbf{R_{oc}}$ is a $3 \times 3$ rotation matrix and $\mathbf{T_{oc}}$ is a $3 \times 1$ translation vector. Both describe orientation and position of the optical coordinate system with respect to the camera coordinate system. The $3 \times 4$ matrix $[\mathbf{R}_{\mathbf{oc}}^{-1}\ \mathbf{T}_{\mathbf{oc}}]$ in Equation (2) is the *transformation matrix* (from the optical coordinate system into the camera coordinate system).

Let $\alpha$ denote the angle between the $Z_o$-axis of the camera coordinate system and the line segment $\overline{\mathbf{OC}}$ (see Figure 2.11). The rotation matrix is

$$\mathbf{R_{oc}} = \begin{bmatrix} \cos(\alpha+\omega) & 0 & -\sin(\alpha+\omega) \\ 0 & 1 & 0 \\ \sin(\alpha+\omega) & 0 & \cos(\alpha+\omega) \end{bmatrix} \tag{3}$$

where $\omega$ is the principal angle of the camera model. The translation vector is

$$\mathbf{T_{oc}} = \begin{bmatrix} R\sin\alpha \\ 0 \\ R\cos\alpha \end{bmatrix} \tag{4}$$

where $R$ is the off-axis distance of the panoramic camera model.

### 2.4.3. ANGULAR IMAGE COORDINATE SYSTEM

Another way of expressing an image point $(x, y)$ is defined by an *angular image coordinate system*. The coordinates are denoted as $(\alpha, \beta)$, where $\alpha$ has been defined in the previous subsection, and $\beta$ is the angle between the $Z_c$-axis of the optical coordinate system and the line incident with the associated optical center and the image point $(x, y)$; see Figure 2.12. The

[6] Read "if and only if".

angular image coordinate system allows some kind of "unification" between the image, camera, and optical coordinate systems by describing an image point $(x, y)$ using the angles $\alpha$ and $\beta$.

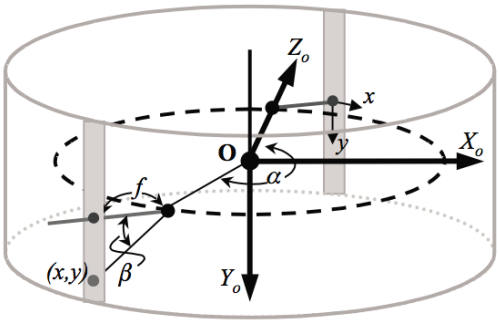



*Figure 2.12.* Definition of angular image coordinates $\alpha$ and $\beta$.

The conversion between image coordinates $(x, y)$ and angular image coordinates $(\alpha, \beta)$ is defined by

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{2\pi x}{W\mu} \\ \arctan\left(\frac{y}{f}\right) \end{bmatrix}$$

where $f$ is the effective focal length of the panoramic camera and $W$ is the number of image columns.

# CAMERA CALIBRATION

*Camera calibration* is the process of determining values of camera parameters. The camera calibration problem is fundamental for many computer vision tasks, robotic applications, or image-based visualizations. Especially, it needs to be solved for those applications which require precise 3D scene reconstructions or measurements based on camera data.

Production-site facilities allow the calibration of manufactured cameras (Klette et al., 2001). This chapter, however, considers the camera calibration problem as it appears at the moment when using a camera. By utilizing geometric relations between known geometric information about the scene or calibration objects (e.g., relative coordinates of 3D points, length of lines) and the projections of geometric features in the acquired images, Theoretically it is in general possible to infer values of camera parameters with high accuracy. Images whose associated values of camera parameters are known are called *calibrated images*.

This chapter provides a comparative discussion of different ways for calibrating panoramic cameras.

## 3.1. Three Methods

Although the general calibration scenario (calibration objects, localization of calibration marks, calculation of internal and external camera parameters, and so fort), and some of the used procedures (e.g., detection of calibration marks) may be similar for both planar and cylindric images, different camera architectures require actually adjusted camera calibration methods. The panoramic camera model contains nonlinear multiple projection centers and a nonplanar image projection surface (e.g., cylindric). This nonlinearity defines new challenges for camera calibration.

Criteria with respect to the performance of camera calibration include the number of images required, the difficulty of preprocessing, computational complexity, dimensionality of parameter estimation, sample data required for calibration, (non-)dependence from initial values in the pa-

rameter estimation process, error sensitivity of parameter estimation with respect to incremental errors, and the reproducible verification of camera parameters.

Preprocesses usually involve the detection of traceable features (e.g., calibration markers, lines) in an image, and they can be more difficult than the camera calibration process itself.

Optimization is often necessary for improving the quality of camera calibration. It is well known that the accuracy and the error sensitivity of estimated parameters critically depends on dimensionality and computational complexity of exploited objective functions. The simplification of objective functions or the reduction of dimensionality (without loss the descriptiveness of objective functions) are key issues for the design of a calibration method. We discuss geometric optimization options, but (in this report) not numerical optimization problems or generic methods.

The chapter presents three different approaches. A special point of interest is how existing concepts, developed for pinhole cameras, can be applied to cylindric panoramic cameras. In particular, how far can we restrict ourself on using just linear geometric features?

At first we present a *point-based approach*, which is commonly used for calibration of planar pinhole cameras. The basic idea is to find the minimum difference between actual and ideal projections of 3D points (i.e., calibration marks) where relative coordinates are known based on measurements. This approach results in a discussion about unstable parameter estimation for panoramic cameras due to non-linearity and high dimensionality. [It requires further research to be (possibly) applicable.]

The second method is an *image correspondence approach*. It is known that (just) image correspondence information is insufficient for planar pinhole camera calibration. We show that this changes for panoramic camera calibration. This approach requires neither scene measures nor any calibration object. Therefore it avoids influence of extrinsic camera parameters on the calibration of intrinsic parameters. [This approach is potentially possible, but requires also further research.]

Finally, we present a *parallel line approach* that uses geometric properties of parallel line segments (calibration lines) available in the scene. In this case, a panoramic camera is well posed if the axis of the image cylinder is parallel to at least three of those straight line segments. This approach allows the most accurate and numerical stable (compared to the other two methods) calibration.

We split the calibration process for a panoramic camera into two steps. The first step is to calibrate the effective focal length $f$ and the *principal row* $v_c$ (defined by the intersection of the cylindric sensor area with the base plane). The second step is to calibrate off-axis distance $R$ and

principal angle $\omega$. This allows to explore linear features independent of non-linear features, to divide dimensionality, and to reduce computational complexity. This two-step calibration process is applicable regardless of the used panoramic camera or progressive scanning approach. The separability of the total calibration process shows that panoramic camera geometry is a mixture of linear and non-linear components.

Section 3.2 discusses the 'linear' step of panoramic camera calibration for $f$ and $v_c$. The rest of the chapter presents the three approaches (listed above) for the second, the 'non-linear' step of calibration for $R$ and $\omega$. The standard approach for pinhole cameras (i.e., the point-based approach) serves as motivation for discussing further improvements. The performance-improved, image-correspondence based approach follows next. Finally, the chapter shows that the use of parallel lines (note: the exploitation of linear features) leads to the most stable solution among these three methods.

## 3.2. Effective Focal Length and Principle Row

We calibrate the effective focal length $f$ (measured in pixels) and the *principle row* $v_c$. Because of the linear nature of a single sensor column, the projective geometry is a 2D simplification of the 3D projective geometry (Tsai, 1987; Faugeras, 1993) of the pinhole camera. Given a calibration object, we calibrate $f$ and $v_c$ by minimizing differences between actual and ideal projections of known 3D points on a calibration object.

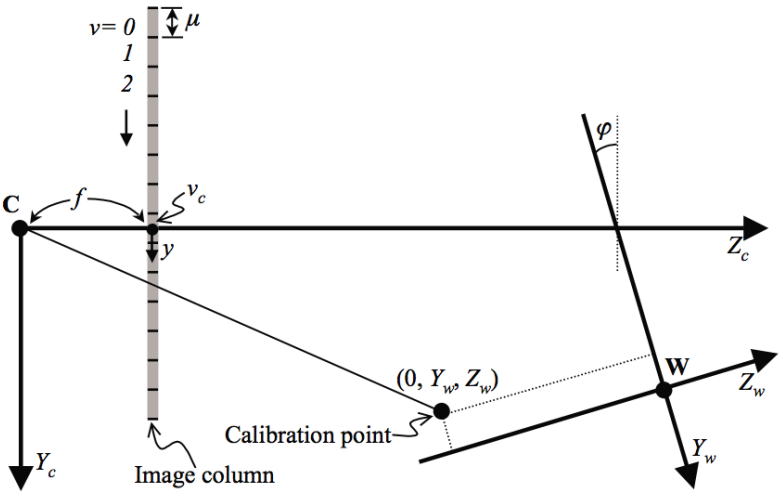



*Figure 3.1.* The first ('linear') step of panoramic camera calibration.

In this case a 2D space is sufficient to describe necessary geometric relations, see Figure 3.1. **C** is the associated optic center. The 2D coordinate system shows projections of 3D calibration points (positioned on a 3D calibration object). The coordinates are denoted by $(Y_c, Z_c)$. (We use $Y$ and $Z$ instead of $X$ and $Y$ in correspondence with the used 3D coordinate system in this report.) The $Y$-axis of the camera coordinate system is defined parallel to the sensor column; the $Z$-axis is perpendicular to the sensor column and passes through it.

Two image coordinate systems are used for describing an image. One is discrete with integer coordinate $v$, and the other one is the Euclidean image coordinate system with real coordinate $y$; see Figure 3.1. The conversion between coordinates $v$ and $y$ is as follows:

$$v = v_c + \frac{y}{\mu}$$

where $v_c$ is the coordinate of that image or sensor point where the $Z$-axis of the camera coordinate system intersects the image, and $\mu$ is the height of a pixel (i.e., pixel in the cell model) in the image. The image point with coordinate $v_c$ is referred to as *principle point* of the sensor or image.

The direction of the $Y$-axis of the camera coordinate system is defined to be the same as that of the $y$-axis in the Euclidean image coordinate system. A point $(Y_c, Z_c)$ is projected onto an image column at position $y$ as follows:

$$y = \frac{fY_c}{Z_c}$$

where $f$ is the effective focal length of the panoramic camera. The relation between pixel $v$ (in the point model) and $(Y_c, Z_c)$ can be described in matrix form as follows:

$$\begin{bmatrix} sv \\ s \end{bmatrix} = \begin{bmatrix} \frac{f}{\mu} & v_c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} Y_c \\ Z_c \end{bmatrix}$$

where $s$ is any scalar. $f_\mu = \frac{f}{\mu}$ is the camera's effective focal length measured in pixels.

The calibration object in Figure 3.1 is (theoretically) a planar region. In the plane of this calibration object we consider a 2D world coordinate system, with its origin at **W**. Any point in the plane of Figure 3.1 has thus world coordinates $(Y_w, Z_w)$.

The acute angle between the $Y_w$-axis and the $Z$-axis (of the camera coordinate system) is denoted as $\varphi$. The origin **C** of the camera coordinate system has world coordinates $(t_y, t_z)$.

A calibration point $(Y_w, Z_w)$ is transformed as follows, from world coordinates into camera coordinates:

$$\begin{bmatrix} Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) & t_y \\ \sin(\varphi) & \cos(\varphi) & t_z \end{bmatrix} \begin{bmatrix} Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Thus, the relation between a calibration point $(Y_w, Z_w)$ and its projection $v$ in the image can be expressed as follows:

$$\begin{aligned} \begin{bmatrix} sv \\ s \end{bmatrix} &= \begin{bmatrix} f_\mu & v_c \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) & t_y \\ \sin(\varphi) & \cos(\varphi) & t_z \end{bmatrix} \begin{bmatrix} Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} f_\mu \cos(\varphi) + v_c \sin(\varphi) & -f_\mu \sin(\varphi) + v_c \cos(\varphi) & f_\mu t_y + v_c t_z \\ \sin(\varphi) & \cos(\varphi) & t_z \end{bmatrix} \begin{bmatrix} Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned}$$

The value of $v$ is equal to the following:

$$\frac{Y_w(f_\mu \cos(\varphi) + v_c \sin(\varphi)) + Z_w(v_c \cos(\varphi) - f_\mu \sin(\varphi)) + (f_\mu t_y + v_c t_z)}{Y_w \sin(\varphi) + Z_w \cos(\varphi) + t_z} \quad (1)$$

The values of $f_\mu$ and $v_c$ can therefore be determined by a given set of calibration points and their corresponding projections. Equation (1) can be rearranged into a linear equation of five unknowns denoted as $X_i$, where $i \in [1, 2, \ldots, 5]$ :

$$Y_w X_1 + Z_w X_2 - v Y_w X_3 - v Z_w X_4 + X_5 = v$$

where

$$\begin{aligned} X_1 &= \frac{f_\mu \cos(\varphi) + v_c \sin(\varphi)}{t_z} \\ X_2 &= \frac{v_c \cos(\varphi) - f_\mu \sin(\varphi)}{t_z} \\ X_3 &= \frac{\sin(\varphi)}{t_z} \\ X_4 &= \frac{\cos(\varphi)}{t_z}\,, \quad \text{and} \\ X_5 &= \frac{(f_\mu t_y + v_c t_z)}{t_z} \end{aligned}$$

It follows that at least five pairs of calibration points and their projections are necessary to determine $f_\mu$ and $v_c$.

Now assume that $n$ such pairs $\{(Y_{wi}, Z_{wi}), v_i\}$ are given, where $n > 5$; $(Y_{wi}, Z_{wi})$ denotes the $i$th calibration point, and $v_i$ denotes the corresponding projection. This allows to solve the following overdetermined system of linear equations:

$$\begin{bmatrix} Y_{w1} & Z_{w1} & -v_1 Y_{w1} & -v_1 Z_{w1} & 1 \\ Y_{w2} & Z_{w2} & -v_2 Y_{w2} & -v_2 Z_{w2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{wn} & Z_{wn} & -v_n Y_{wn} & -v_n Z_{wn} & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

After having a least-square solution of $(X_1, X_2, X_3, X_4, X_5)^{\mathrm{T}}$, the values of $f_\mu$ and $v_c$ can be calculated as follows:

$$\begin{bmatrix} f_\mu \\ v_c \end{bmatrix} = \begin{bmatrix} X_4 & X_3 \\ -X_3 & X_4 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

This completes the first step of the calibration of the panoramic camera, which provides the effective focal length $f$ (measured in pixels) and the principle row $v_c$.

The following sections present and characterize different options for a second step of camera calibration. Hereafter we can always assume that effective focal length $f_\mu$ and principle row $v_c$ are known.

## 3.3. Point-based Approach

This is a first option for calibrating off-axis distance $R$ and principal angle $\omega$. A common (and straightforward) approach 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. (Such a point-based approach was already used in Section 3.2 for the first step of the calibration process.)

We start with deriving the general projection formula for cylindric panoramic cameras, then we calibrate $R$ and $\omega$ using this formula, and finally we discuss problems and difficulties of this approach with respect to performance criteria.

### 3.3.1. PROJECTION FORMULA

Consider a known 3D point $\mathbf{P}$ with coordinates $(X_w, Y_w, Z_w)$ in a world coordinate system. Point $\mathbf{P}$ is first transformed into the camera coordinate system before calculating its projection in the image. We denote the coordinates of $\mathbf{P}$ with respect to the camera coordinate system as $(X_o, Y_o, Z_o)$.

We have

$$\begin{bmatrix} X_o \\ Y_o \\ Z_o \end{bmatrix} = [\mathbf{R_{wo}} \; -\mathbf{R_{wo}}\mathbf{T_{wo}}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

where $\mathbf{R_{wo}}$ is a $3 \times 3$ rotation matrix and $\mathbf{T_{wo}}$ is a $3 \times 1$ translation vector. Both together describe the affine transform of the world coordinate system into the camera coordinate system. The $3 \times 4$ matrix $[\mathbf{R_{wo}} \; -\mathbf{R_{wo}}\mathbf{T_{wo}}]$ is the *transformation matrix*, and its twelve parameters are denoted as follows:

$$[\mathbf{R_{wo}} \; -\mathbf{R_{wo}}\mathbf{T_{wo}}] = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix}$$

The projection of $(X_o, Y_o, Z_o)$ can be expressed in image coordinates $(u, v)$, and $u$ and $v$ can be determined separately.

Let $\alpha$ be the angle between the $Z$-axis of the panorama coordinate system and the line segment $\overline{\mathbf{OC}}$, where $\mathbf{C}$ is the focal point associated to image column $u$; see Figure 3.2(B). We have

$$x = \frac{\alpha \mu W}{2\pi}$$

where $x$ is a Euclidean (i.e., real) image coordinate. Thus we have

$$u = \frac{\alpha W}{2\pi} \tag{2}$$

We still have to determine $\alpha$.

Consider a 3D point $\mathbf{P}$ with coordinates $(X_o, Y_o, Z_o)$, and let $\mathbf{Q}$ be its projection onto the $XZ$-plane of the camera coordinate system; see Figure 3.2(A). Point $\mathbf{Q}$ has camera coordinates $(X_o, 0, Z_o)$.

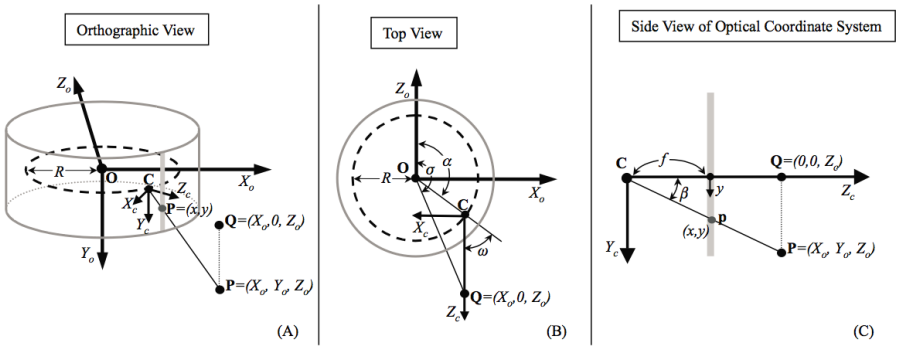



*Figure 3.2.* Projection geometry of a cylindric panorama.

From the top-view of the projection geometry [see Figure 3.2(B)] we conclude that

$$\alpha = \sigma - \angle\mathbf{COQ} = \sigma - \omega + \angle\mathbf{CQO}$$

Thus we have

$$\alpha = \begin{cases} \arctan\left(\frac{X_o}{Z_o}\right) - \omega + \arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right) & \text{if } Z_o \geq 0 \\ \pi + \arctan\left(\frac{X_o}{Z_o}\right) - \omega + \arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right) & \text{otherwise} \end{cases}$$

If $\alpha$ exceeds $2\pi$ then it is transformed modulo $2\pi$ into $[0, 2\pi)$. For the case of interest (i.e., for $Z_o \geq 0$) we can simplify the equation as follows:

$$\begin{aligned} \alpha &= \arctan\left(\frac{X_o}{Z_o}\right) - \omega + \arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right) \\ &= \arcsin\left(\frac{X_o}{\sqrt{X_o^2+Z_o^2}}\right) - \omega + \arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right) \\ &= \arcsin\left(\frac{X_o}{\sqrt{X_o^2+Z_o^2}}\sqrt{1-\frac{R^2\sin^2\omega}{X_o^2+Z_o^2}} + \frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\sqrt{1-\frac{X_o^2}{X_o^2+Z_o^2}}\right) - \omega \\ &= \arcsin\left(\frac{X_o\left(\sqrt{X_o^2+Z_o^2-R^2\sin^2\omega}\right) + Z_o R\sin\omega}{X_o^2+Z_o^2}\right) - \omega. \end{aligned} \tag{3}$$

This allows to calculate $u$.

Finally, for the calculation of $v$ we can make use of the angular coordinate $\beta$ as shown in Figure 3.2(C). This is the angle between the $Z$-axis of the optical coordinate system, with origin at $\mathbf{C}$, and the line segment $\overline{\mathbf{CP}}$. We have that

$$y = f\tan\beta$$

where $y$ is a Euclidean (i.e., real) image coordinate, and thus

$$\begin{aligned} v &= \frac{f\tan\beta}{\mu} + v_c \\ &= f_\mu\tan\beta + v_c \end{aligned} \tag{4}$$

where $f_\mu$ is the camera's effective focal length measured in pixels, and $v_c$ is the principle row.

Points $\mathbf{P}$ and $\mathbf{Q}$ have coordinates $(0, Y_c, Z_c)$ and $(0, 0, Z_c)$ with respect to the optical coordinate system, respectively, where $Y_c = Y_o$. From the side

view of the optical coordinate system, with origin at **C** [see Figure 3.2(C)], the angular coordinate $\beta$ can be calculated by

$$\beta = \arctan\left(\frac{Y_c}{Z_c}\right)$$

where

$$Z_c = \frac{\overline{\mathbf{OQ}}\sin(\angle\mathbf{COQ})}{\sin\omega}$$

Thus, we have

$$\begin{aligned}
\beta &= \arctan\left(\frac{Y_o\sin\omega}{\sqrt{X_o^2+Z_o^2}\sin\left(\omega-\arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right)\right)}\right) \\
&= \arctan\left(\frac{Y_o\sin\omega}{\sqrt{X_o^2+Z_o^2}\left(\sin\omega\cos\arcsin\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right)-\cos\omega\left(\frac{R\sin\omega}{\sqrt{X_o^2+Z_o^2}}\right)\right)}\right) \\
&= \arctan\left(\frac{Y_o}{\sqrt{X_o^2+Z_o^2}\left(\sqrt{1-\frac{R^2\sin^2\omega}{X_o^2+Z_o^2}}-\cos\omega\frac{R}{\sqrt{X_o^2+Z_o^2}}\right)}\right) \\
&= \arctan\left(\frac{Y_o}{\sqrt{X_o^2+Z_o^2-R^2\sin^2\omega}-R\cos\omega}\right) \qquad (5)
\end{aligned}$$

and this concludes the calculation of $v$.

### 3.3.2. CALIBRATION OF $R$ AND $\omega$

We calibrate the values of the off-axis distance $R$ and the principal angle $\omega$. Assume $n$ pairs of known 3D points $(X_{wi}, Y_{wi}, Z_{wi})$ (in world coordinates) and their actual projections $(\hat{u_i}, \hat{v_i})$ (in image coordinates). The use of a hat '$\hat{}$' indicates that this parameter may be inaccurate.

In contrast, let $(u_i, v_i)$ be the ideal projection of the 3D point $(X_{wi}, Y_{wi}, Z_{wi})$. We want to minimize the following error:

$$\sum_{i=1}^{n}(\hat{u_i}-u_i)^2+(\hat{v_i}-v_i)^2$$

where the value of $u_i$ can be obtained from Equations (2) and (3). The value of $v_i$ can be obtained from Equations (4) and (5). After some minor algebraic transformations, this problem proves to be equivalent to a

minimization of the following:

$$\sum_{i=1}^{n}\left(\sin\left(\frac{2\hat{u}_i\pi}{W}+\omega\right)-\frac{X_{oi}A+Z_{oi}R\sin\omega}{X_{oi}^2+Z_{oi}^2}\right)^2+\left(\hat{v}_i-\frac{f_\mu Y_{oi}}{A-R\cos\omega}+v_c\right)^2 \quad (6)$$

where $A=\sqrt{X_{oi}^2+Z_{oi}^2-R^2\sin^2\omega}$ and

$$\begin{bmatrix} X_{oi} \\ Y_{oi} \\ Z_{oi} \end{bmatrix} = \begin{bmatrix} X_{wi}t_{11}+Y_{wi}t_{12}+Z_{wi}t_{13}+t_{14} \\ X_{wi}t_{21}+Y_{wi}t_{22}+Z_{wi}t_{23}+t_{24} \\ X_{wi}t_{31}+Y_{wi}t_{32}+Z_{wi}t_{33}+t_{34} \end{bmatrix}$$

Parameters $f_\mu$ and $v_c$ are assumed to be pre-calibrated and (thus) known. Therefore, there are 14 parameters in total here to be estimated using a nonlinear least-square optimization method as described (e.g.) in (Gill et al., 1981). These 14 parameters contain the parameters $R$, $\omega$ as well as the twelve (intermediate) unknowns of the transformation matrix.

### 3.3.3. DISCUSSION

The error function of Equation (6) is rather complicated. The parameters to be estimated are enclosed in sine functions, and square roots are involved in both numerator and denominator of the fractions. The dimensionality is high due to the fact that the extrinsic parameters in $\mathbf{R_{wo}}$ and $\mathbf{T_{wo}}$ are unavoidable in this approach. Thus, a 'large' set of 3D points is needed for reasonably accurate estimation.

Human intervention might be required for this calibration approach for identifying the projections of those 3D points in a real scene used as calibration marks. If a specially designed calibration object is used, this process can

| Approaches | Number of Image Required | Pre-process | Complexity | Dimensionality | Initial Value Dependence | Error Sensitivity | Estimated $R$ | Estimated $\omega$ |
|---|---|---|---|---|---|---|---|---|
| **Point-based** | one | identifying projections of 3D points | nonlinear form e.g. sine, fractions and square roots | 14 unknowns [R\|T]* unavoidable | high | exponential growing | actual | actual |
| **Image Correspondence** | two | searching image corresponding points | linear form | four unknowns [R\|T]* not required | ignorable | exponential growing | ratio | actual |
| **Parallel-line-based** | one | identifying projections of lines | linear form | three unknowns [R\|T]* avoidable | ignorable | linear growing | actual | actual |

* [R|T] is a matrix that characterizes camera extrinsic parameters, where **R** is a rotation matrix and **T** is a translation vector.

*Table 3.1:* Summary for the performance of different camera calibration approaches.

be supported by an automatic calibration mark detection program, where marks are located with sub-pixel accuracy (using, e.g., centroid calculation within a mark's region).

Following this approach, the accuracy of the calibration result critically depends on the accuracy of initial values (i.e., 3D points and identified projections). Error sensitivity analysis shows exponential growth of resulting errors. The high dimensionality of the error function is a critical issue of the point-based approach, and a substantial reduction of the exponential growth of errors in relation to errors of initial values remains an open problem.

Assessments are summarized in Table 3.1, which already shows assessments for two more approaches still to be presented in the next sections of this chapter. These two approaches do not require the camera's extrinsic parameters for calibrating $R$ and $\omega$. At first (in the next section) we discuss the possibility of calibration from image correspondences. For this approach we simply assume two concentric panoramic images whose associated base-circle centers $\mathbf{O}$ are identical, as well as their rotational axes. The search for corresponding points in both images is of the same complexity as in a pinhole camera case, with robust automated solutions if a specially designed calibration object is used.

### 3.4. Image Correspondence Approach

This section discusses the possibility of calibrating $R$ and $\omega$ by using corresponding image points in two uncalibrated concentric panoramas. This approach requires neither scene measures nor any calibration object and is (thus) independent from the camera's extrinsic parameters.

The basic idea of this approach is similar to a recovery of epipolar geometry: use equations of epipolar curves (which are just straight lines for a pinhole camera but more complex for polycentric panoramas; see next chapter) for linking provided corresponding points; then calibrate parameters by optimization. The first step is known as fundamental (or essential) matrix estimation in case of pinhole cameras (see, e.g., (Liu et al., 2005) for recent studies in case of pinhole cameras).

Consequently, this section makes use of a result reported in (Huang et al., 2001b): the analytic representation of epipolar curves for concentric panoramic images. Other categories of polycentric panoramas could be used as well for discussing an image correspondence approach for calibrating $R$ and $\omega$. However, a discussion of concentric panoramas should suffice for the purpose of explaining possible approaches in this report.

### 3.4.1. CONCENTRIC PANORAMIC PAIR

We assume concentric panoramas where the effective focal length and the angular unit are constant for all images in the considered class (of two images). A concentric panoramic pair can be acquired in various ways, by using different or identical off-axis distances, or different or identical principal angles. We discuss the three possible options and will outline which option gives the best calibration performance. This subsection derives the error function for the defined calibration approach in case of concentric panoramas; and this already specifies the calibration procedure.

Let $(x_1, y_1)$ and $(x_2, y_2)$ be a pair of corresponding image points in a concentric pair of panoramas $E_{\mathcal{P}_1}(R_1, f, \omega, \gamma)$ and $E_{\mathcal{P}_2}(R_2, f, \omega, \gamma)$. Using the general epipolar curve equation provided in the next chapter we obtain that both points have to satisfy the following epipolar curve equation for concentric panoramas (also assuming identical effective focal lengths and identical angular units):

$$y_2 = y_1 \cdot \frac{R_2 \sin\omega - R_1 \sin(\alpha_2 + \omega - \alpha_1)}{R_2 \sin(\alpha_1 + \omega - \alpha_2) - R_1 \sin\omega} \tag{7}$$

where $\alpha_1 = \frac{2\pi x_1}{\mu W_1}$ and $\alpha_2 = \frac{2\pi x_2}{\mu W_2}$. The equation can be rearranged as follows:

$$y_2 R_2 \sin((\alpha_1 - \alpha_2) + \omega) - y_2 R_1 \sin\omega + y_1 R_1 \sin((\alpha_2 - \alpha_1) + \omega) - y_1 R_2 \sin\omega = 0$$

Let $(\alpha_1 - \alpha_2) = \sigma$. We have

$$\begin{aligned} y_2 \sin\sigma R_2 \cos\omega + y_2 \cos\sigma R_2 \sin\omega - y_1 \sin\sigma R_1 \cos\omega \\ + y_1 \cos\sigma R_1 \sin\omega - y_2 R_1 \sin\omega - y_1 R_2 \sin\omega \;=\; 0 \end{aligned}$$

Furthermore, we obtain

$$\begin{aligned} y_2 \sin\sigma R_2 \cos\omega + (y_2 \cos\sigma - y_1) R_2 \sin\omega \\ - y_1 \sin\sigma R_1 \cos\omega + (y_1 \cos\sigma - y_2) R_1 \sin\omega \;=\; 0 \end{aligned}$$

We observe from those equations that the ratio $R_1 : R_2$ and the value of $\omega$ can be calibrated. The values of $R_1$ and $R_2$ are not computable if using this approach alone.

Now assume $n$ pairs of corresponding image points $(x_{1i}, y_{1i})$ and $(x_{2i}, y_{2i})$ in both panoramas, where $i \in [1..n]$. The values of corresponding indices $\sigma_i$ are assumed to be be calculated in advance. We use the optimization method of sequential quadratic programming (Gill et al., 1981) for estimating $\frac{R_1}{R_2}$ and $\omega$. The error function to be minimized is the following:

$$\sum_{i=1}^{n} [y_{i2} \sin\alpha_i X_1 + (y_{i2} \cos\alpha_i - y_{i1}) X_2 - y_{i1} \sin\alpha_i X_3 + (y_{i1} \cos\alpha_i - y_{i2}) X_4]^2$$

subject to the equality constraint $X_1X_4 = X_2X_3$, where $X_1 = R_2 \cos\omega$, $X_2 = R_2 \sin\omega$, $X_3 = R_1 \cos\omega$, and $X_4 = R_1 \sin\omega$.

Once the values of $X_1$, $X_2$, $X_3$, and $X_4$ are obtained from the specified minimization, $\frac{R_1}{R_2}$ and $\omega$ can be calculated by

$$\frac{R_1}{R_2} = \frac{\sqrt{X_3^2 + X_4^2}}{\sqrt{X_1^2 + X_2^2}}$$

and

$$\omega = \arccos\left(\frac{X_1}{\sqrt{X_1^2 + X_2^2}}\right)$$

### 3.4.2. EXPERIMENTAL RESULTS

The error function of this approach is basically in linear form, and there are only four unknowns to be estimated, namely $X_1$, $X_2$, $X_3$, and $X_4$. This means that at least four pairs of corresponding image points (in general positions) are required for solving this minimization problem uniquely. This can be considered as a great improvement compared to the point-based approach.

However, experiments indicate inaccuracies; estimated values for concentric panoramas of real scenes appeared to be erroneous compared to the known (i.e., pre-calibrated) parameter values. Such an experiment using

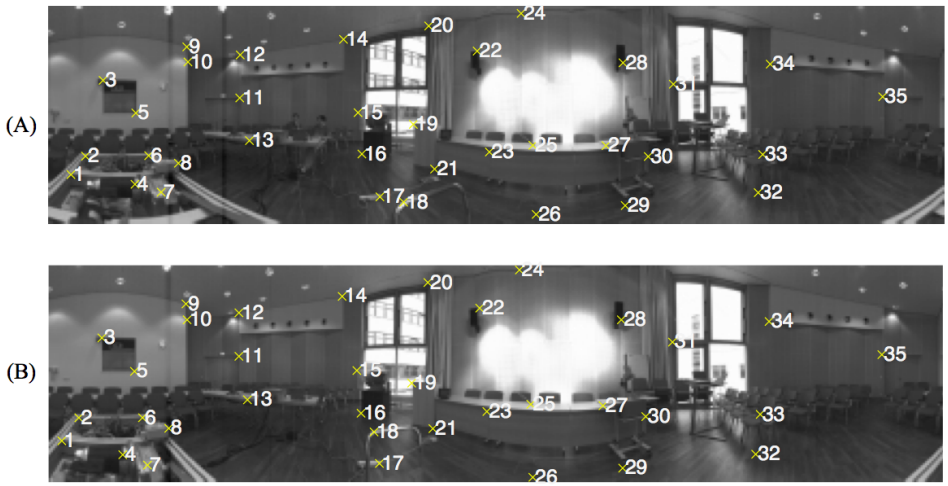



*Figure 3.3.* Two concentric panoramas (showing a seminar room at DLR in Berlin-Adlershof) with 35 pairs of corresponding image points, used for the calibration of off-axis distance $R$ and principal angle $\omega$.

| Input data error in pixel | Estimated $\omega$ error in (%) |
|---|---|
| 0.0 | 0.00 |
| 0.5 | 0.17 |
| 1.0 | 0.52 |
| 1.5 | 2.78 |
| 2.0 | 10.34 |
| 3.0 | 28.14 |
| 4.0 | 68.25 |
| 5.0 | 155.62 |

*Table 3.2:* Error sensitivity results of calibration by image correspondence.

real scene data is illustrated in Figure 3.3. In this example there are 35 pairs of corresponding image points identified manually, marked by crosses and indexed by numbers. We skip details because results also depend on the accuracy of the pre-calibration method.

The error sensitivity of the approach can be analyzed in an objective way by a simulation on synthetic data. Ground-truth data can be generated by attempting a simulation of real scenes (e.g., as shown in Figure 3.3), and errors are then simulated by additive random noise with normal distribution, perturbating the coordinates of ideal pairs of corresponding image points. A few calibration results for $\omega$ in dependence of additive errors are shown in Table 3.2 and illustrate that the estimated result is rather sensitive to these errors. In general, errors of estimated parameters increased exponentially with respect to errors in input data.

A reason why this image correspondence approach is sensitive to input error is that the values of the coefficients in the error function are likely to be numerically very close to one-another, and these values depend on the selected pairs of corresponding points. Possible ways for suppressing the error-sensitiveness [without relying on the 'robustness' of numerical methods as, for example, discussed in (Gill et al., 1981)] include

(1) increase the number of pairs of corresponding points,
(2) place calibration object closer to the camera (i.e., larger disparities),
(3) design a special calibration object that allows that all coded points on the calibration-object appear twice in a single panorama.

Proposal (1) seemed (in experiments) to allow only minor improvements.

Despite the error problem, this approach is unable to recover the absolute value of $R$, only relative values as said above. Assessments for this approach are summarized in Table 3.1. The drawbacks of the two approaches discussed so far are a motivation to search for an approach where linear geometric relations between 3D scenes panoramic images can be

utilized. The next section investigates a parallel-line-based approach, which allows a further reduction of the dimensionality of the error function, simplifies the computational complexity by using linear geometric features only, and proved to be less sensitive to errors in comparison to the first two approaches (only as specified above; further refinements might be able to change these relations).

## 3.5. Parallel-line-based Approach

This section calibrates off-axis distance $R$ and principal angle $\omega$ using an approach that has been widely used for pinhole camera calibration: we call it the parallel-line-based approach.

The general intention is to find a single linear equation that links 3D geometric scene features to the image cylinder such that (by providing sufficient scene measurements) we are able to calibrate $R$ and $\omega$ with high accuracy. Usable 3D scene features are, for example, distances, lengths, or angular configurations of straight line segments. This section informs about two alternatives for such an approach, one only based on parallel line segments, and another one which also takes orthogonal configurations of straight lines into consideration.

### 3.5.1. ASSUMPTIONS

We assume there are at least three straight line segments in the captured real scene (e.g., straight edges of doors or windows), which are parallel to the axis of the associated image cylinder. [The latter assumption is normally satisfiable by using a "bulls eye" or a more advanced leveling device.]

For each straight line segment we assume that both end points are visible (from the camera) and identifiable in the panoramic image, and that we have an accurate measurement of the distance between these two end points. The projected line segment (in the panoramic image) should ideally be in a single image column, and we will assume this. [The length of a projected line segment in arbitrary position in the panoramic image, forming a curved line, could be measured with methods as described in (Klette and Rosenfeld, 2004).]

Furthermore, for each of the selected straight line segments we assume either (for the first alternative) that there is a second useable straight line segments in the scene which is parallel to the first, and where the distance between both lines was also accurately measurable, or (for the second alternative) that there exist two more useable parallel straight line segments such that all three segments form an orthogonal configuration.

Effective focal length $f$ (in pixels) and principle row $v_c$ are again assumed to be already known (from the first step of the calibration procedure).

### 3.5.2. FIRST ALTERNATIVE (DISTANCE CONSTRAINT)

Any useable straight line segment in the 3D scene is denoted as $\mathcal{L}$ and indexed where needed for the distinction of multiple lines. The (Euclidean) distance of two visible points on a line $\mathcal{L}$ is denoted as $H$ (like 'height'). The length of a projection of a line segment on an image column $u$ is denoted as $h$ and measured in pixels. Examples of $H_i$ and corresponding $h_i$ values are illustrated in Figure 3.4(A), where $i \in [1, \ldots, 5]$.

The distance $D_{ij}$ between two parallel lines $\mathcal{L}_i$ and $\mathcal{L}_j$ is the length of a line segment that connects both and is perpendicular to them. If the distance between two straight line segments is available then we say that both lines form a *pair of lines*. A line segment may be paired up with more than just one other line segment. Figure 3.4(A) shows three pairs of lines, namely ($\mathcal{L}_1$, $\mathcal{L}_2$), ($\mathcal{L}_3$, $\mathcal{L}_4$), and ($\mathcal{L}_4$, $\mathcal{L}_5$).

Consider two straight segments $\mathcal{L}_i$ and $\mathcal{L}_j$ in 3D space and the image columns of their projections, denoted as $u_i$ and $u_j$, respectively. The optical centers associated to image columns $u_i$ and $u_j$ are denoted as $\mathbf{C}_i$ and $\mathbf{C}_j$ respectively. The distance of the two associated image columns is $d_{ij} =$

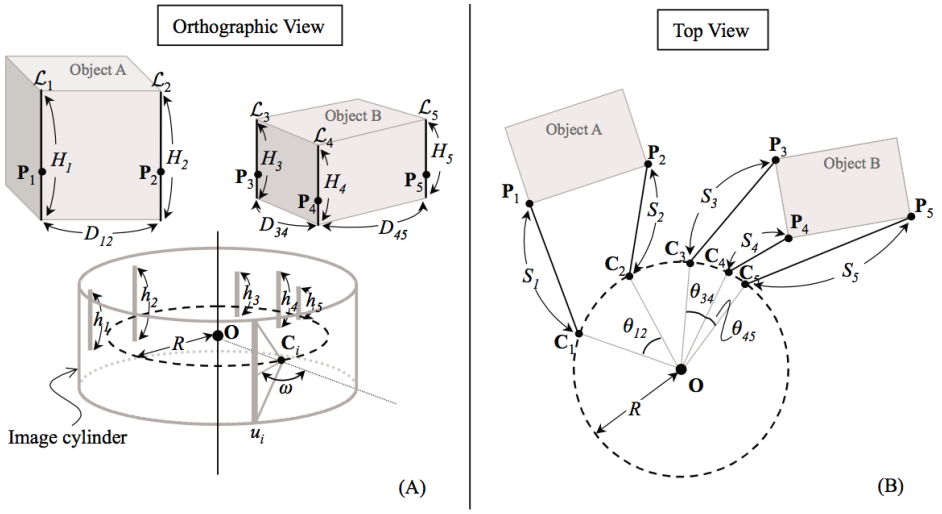



*Figure 3.4.* Configurations of parallel straight lines in the 3D scene and on the image cylinder. (A) Orthographic side view. (B) Top view.

$|u_i - u_j|$ (in pixels). The angular distance of two image columns, associated to line segments $\mathcal{L}_i$ and $\mathcal{L}_j$, is the angle between line segments $\overline{\mathbf{C}_i\mathbf{O}}$ and $\overline{\mathbf{C}_j\mathbf{O}}$, where $\mathbf{O}$ is the center of the base circle. We denote the angular distance of a pair $(\mathcal{L}_i, \mathcal{L}_j)$ of lines as $\theta_{ij}$. Examples of angular distances for two pairs of lines are given in Figure 3.4(B). The angular distance $\theta_{ij}$ and $d_{ij}$ are related by

$$\theta_{ij} = \frac{2\pi d_{ij}}{W}$$

where $W$ is the width of the panorama in pixels.

The distance $S$ between a line segment $\mathcal{L}$ and the associated optical center (which 'sees' this line segment) is defined by the length of the line segment starting at the optical center, ending on $\mathcal{L}$ and being perpendicular to $\mathcal{L}$. We have

$$S = \frac{f_\mu H}{h},$$

where $f_\mu$ is the pre-calibrated effective focal length of the camera.

*Geometric Relation*

Now we are ready to formulate a distance constraint by combining all the previously described geometric information. A 2D coordinate system is defined on the base plane for every pair of lines $(\mathcal{L}_i, \mathcal{L}_j)$; see Figure 3.5. Note that even though all the measurements are defined in 3D space, the geometric relation of interest can be described in a 2D space since all the straight segments are assumed to be parallel to the axis of the image cylinder. The origin of the coordinate system is $\mathbf{O}$, and the $Z$-axis is incident with the camera focal point $\mathbf{C}_i$. The $X$-axis is orthogonal to the $Z$-axis and is incident with the base plane. (This coordinate system coincides with the camera coordinate system previously defined but without $Y$-axis.) Such a coordinate system is defined for each pair of lines.

The position of $\mathbf{C}_i$ can now be described by coordinates $(0, R)$, and the position $\mathbf{C}_j$ can be described by coordinates $(R\sin\theta_{ij}, R\cos\theta_{ij})$. The intersection point of line $\mathcal{L}_i$ with the base plane, denoted as $\mathbf{P}_i$, can be expressed by a sum of vector $\overrightarrow{\mathbf{OC}_i}$ and vector $\overrightarrow{\mathbf{C}_i\mathbf{P}_i}$. Thus, we have the following:

$$\mathbf{P}_i = \begin{bmatrix} S_i \sin\omega \\ R + S_i \cos\omega \end{bmatrix}$$

Analogously, the intersection point of line $\mathcal{L}_j$ with the base plane, denoted as $\mathbf{P}_j$, can be described by a sum of vectors $\overrightarrow{\mathbf{OC}_j}$ and $\overrightarrow{\mathbf{C}_j\mathbf{P}_j}$. We have the following:

$$\mathbf{P}_j = \begin{bmatrix} R\sin\theta_{ij} + S_j \sin(\theta_{ij} + \omega) \\ R\cos\theta_{ij} + S_j \cos(\theta_{ij} + \omega) \end{bmatrix}$$

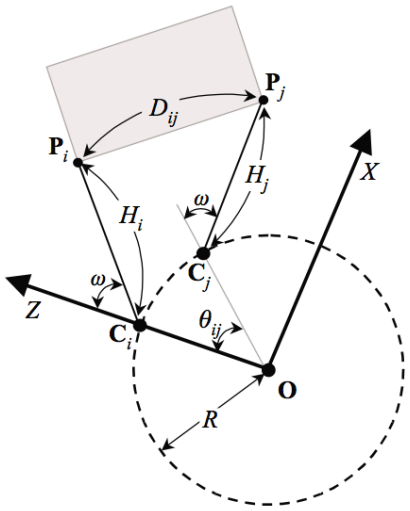



*Figure 3.5.* Coordinate system of a pair of lines.

The distance $D_{ij}$ between points $\mathbf{P}_i$ and $\mathbf{P}_j$ has been measured. We have the following equation:

$$\begin{aligned} D_{ij}^2 &= (S_i \sin\omega - R\sin\theta_{ij} - S_j \sin(\omega+\theta_{ij}))^2 \\ &+ (R + S_i\cos\omega - R\cos\theta_{ij} - S_j\cos(\omega+\theta_{ij}))^2 \end{aligned}$$

This equation can be expanded and rearranged as follows:

$$\begin{aligned} D_{ij}^2 &= S_i^2\sin^2\omega + R^2\sin^2\theta_{ij} + S_j^2\sin^2(\omega+\theta_{ij}) - 2S_iR\sin\omega\sin\theta_{ij} \\ &\quad -2S_iS_j\sin\omega\sin(\omega+\theta_{ij}) + 2RS_j\sin\theta_{ij}\sin(\omega+\theta_{ij}) \\ &\quad +R^2 + S_i^2\cos^2\omega + R^2\cos^2\theta_{ij} + S_j^2\cos^2(\omega+\theta_{ij}) \\ &\quad +2RS_i\cos\omega - 2R^2\cos\theta_{ij} - 2RS_j\cos(\omega+\theta_{ij}) - 2S_iR\cos\omega\cos\theta_{ij} \\ &\quad -2S_iS_j\cos\omega\cos(\omega+\theta_{ij}) + 2RS_j\cos\theta_{ij}\cos(\omega+\theta_{ij}) \\ &= S_i^2 + 2R^2 + S_j^2 + 2RS_i\cos\omega - 2R^2\cos\theta_{ij} \\ &\quad -2S_jR\cos(\omega+\theta_{ij}) - 2S_iR(\sin\omega\sin\theta_{ij} + \cos\omega\cos\theta_{ij}) \\ &\quad -2S_iS_j(\sin\omega\sin(\omega+\theta_{ij}) + \cos\omega\cos(\omega+\theta_{ij})) \\ &\quad +2S_jR(\sin\theta_{ij}\sin(\omega+\theta_{ij}) + \cos\theta_{ij}\cos(\omega+\theta_{ij})) \\ &= S_i^2 + S_j^2 + 2R^2(1-\cos\theta_{ij}) + 2S_iR\cos\omega \\ &\quad -2S_jR(\cos\omega\cos\theta_{ij} - \sin\omega\sin\theta_{ij}) - 2S_iR(\sin\omega\sin\theta_{ij} + \cos\omega\cos\theta_{ij}) \\ &\quad -2S_iS_j\cos\theta_{ij} + 2S_jR\cos\omega \\ &= S_i^2 + S_j^2 + 2R^2(1-\cos\theta_{ij}) + 2(S_i+S_j)R\cos\omega \\ &\quad -2(S_i+S_j)R\cos\omega\cos\theta_{ij} - 2(S_i-S_j)R\sin\omega\sin\theta_{ij} \end{aligned}$$

$$-2S_iS_j\cos\theta_{ij}$$

Finally, we obtain

$$\begin{aligned} 0 &= (1-\cos\theta_{ij})R^2 \\ &+ (S_i+S_j)(1-\cos\theta_{ij})R\cos\omega \\ &- (S_i-S_j)\sin\theta_{ij}R\sin\omega \\ &+ \frac{S_i^2+S_j^2-D_{ij}^2}{2} - S_iS_j\cos\theta_{ij} \end{aligned} \tag{8}$$

*Error Function*
Basically we use Equation (8) as error function. The values of $S_i$, $S_j$, $D_{ij}$, and $\theta_{ij}$ are known. Thus Equation (8) can be arranged into the following linear form

$$K_1X_1 + K_2X_2 + K_3X_3 + K_4 = 0$$

with coefficients $K_i$, $i = 1, 2, 3, 4$, defined as follows:

$$\begin{aligned} K_1 &= 1-\cos\theta_{ij} \\ K_2 &= (S_i+S_j)(1-\cos\theta_{ij}) \\ K_3 &= -(S_i-S_j)\sin\theta_{ij} \\ K_4 &= \frac{S_i^2+S_j^2-D_{ij}^2}{2} - S_iS_j\cos\theta_{ij} \end{aligned}$$

For the three linearly independent variables $X_i$, $i = 1, 2, 3$, we have

$$\begin{aligned} X_1 &= R^2 \\ X_2 &= R\cos\omega \\ X_3 &= R\sin\omega \end{aligned}$$

In this case we can solve for absolute (not just relative) values $R$ and $\omega$ by using all three equations. (If more than three equations are provided then it is possible to apply a linear least-square technique.) The values of $R$ and $\omega$ may be calculated by

$$R = \sqrt{X_1} = \sqrt{X_2^2+X_3^2}$$

and

$$\omega = \arccos\left(\frac{X_2}{\sqrt{X_1}}\right) = \arcsin\left(\frac{X_3}{\sqrt{X_1}}\right) = \arccos\left(\frac{X_2}{\sqrt{X_2^2+X_3^2}}\right)$$

The given dependencies among variables $X_1$, $X_2$, and $X_3$ define multiple solutions of $R$ and $\omega$. To tackle this multiple-solution problem, we constrain the parameter estimation process further by

$$X_1^2 = X_2^2 + X_3^2$$

which is valid because of

$$R^2 = (R\cos\omega)^2 + (R\sin\omega)^2$$

Assume that $n$ copies of Equation (8) are given. We want to minimize the following:

$$\sum_{i=1}^{n} (K_{1n}X_1 + K_{2n}X_2 + K_{3n}X_3 + K_{4n})^2 \tag{9}$$

subject to the equality constraint $X_1 = X_2^2 + X_3^2$, where the values of $K_{in}, i = 1, 2, 3, 4$, are calculated based on measurements in the real scene and in the image. We also use $X_1 = R^2$, $X_2 = R\cos\omega$, and $X_3 = R\sin\omega$. Now, the values of $R$ and $\omega$ can be uniquely (!) calculated as

$$R = \sqrt{X_1}$$

and

$$\omega = \arccos\left(\frac{X_2}{\sqrt{X_1}}\right)$$

Note that even though the additional constraint forces a use of a non-linear optimization method, the accuracy of the method remains at the quality level of a linear parameter estimation procedure.

### 3.5.3. SECOND ALTERNATIVE (ORTHOGONALITY CONSTRAINT)

We say that three parallel line segments $\mathcal{L}_i$, $\mathcal{L}_j$, and $\mathcal{L}_k$ are *orthogonal* iff the plane incident with $\mathcal{L}_i$ and $\mathcal{L}_j$ and the plane incident with $\mathcal{L}_j$ and $\mathcal{L}_k$ are orthogonal. (It follows that $\mathcal{L}_j$ is the intersection of both planes.) For example, line segments $\mathcal{L}_3$, $\mathcal{L}_4$, and $\mathcal{L}_5$ in Figure 3.4(A) are orthogonal.

Consider three orthogonal line segments $\mathcal{L}_i$, $\mathcal{L}_j$, and $\mathcal{L}_k$ in 3D space. We have the measured values of $S_i$, $S_j$, $S_k$, $\theta_{ij}$, and $\theta_{jk}$, obtained the same way as in case of the distance constraint. We define a 2D coordinate system for each group of three orthogonal line segments; see Figure 3.6 for segments ($\mathcal{L}_i$, $\mathcal{L}_j$, $\mathcal{L}_k$).

The position of $\mathbf{C}_j$ is given by coordinates $(0, R)$, the position of $\mathbf{C}_i$ by coordinates $(-R\sin\theta_{ij}, R\cos\theta_{ij})$, and the position of $\mathbf{C}_k$ by coordinates $(R\sin\theta_{jk}, R\cos\theta_{jk})$. Intersection points of line segments $\mathcal{L}_i$, $\mathcal{L}_j$, and $\mathcal{L}_k$ (or

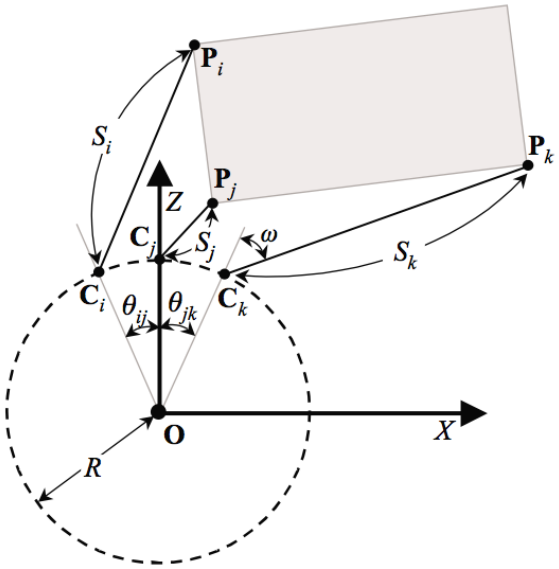



*Figure 3.6.* Coordinate system for three orthogonal line segments.

their linear extensions) with the base-plane are denoted as $\mathbf{P}_i$, $\mathbf{P}_j$, and $\mathbf{P}_k$, respectively. We have

$$\mathbf{P}_i = \begin{bmatrix} -R\sin\theta_{ij} + S_i\sin(\omega - \theta_{ij}) \\ R\cos\theta_{ij} + S_j\cos(\omega - \theta_{ij}) \end{bmatrix}$$

$$\mathbf{P}_j = \begin{bmatrix} S_j\sin\omega \\ R + S_j\cos\omega \end{bmatrix}$$

and

$$\mathbf{P}_k = \begin{bmatrix} R\sin\theta_{jk} + S_k\sin(\theta_{jk} + \omega) \\ R\cos\theta_{jk} + S_k\cos(\theta_{jk} + \omega) \end{bmatrix}$$

Vectors $\overrightarrow{\mathbf{P}_i\mathbf{P}_j}$ and $\overrightarrow{\mathbf{P}_j\mathbf{P}_k}$ are orthogonal; thus we have the following equation:

$$\begin{aligned} 0 \;=\; & (-R\sin\theta_{ij} + S_i\sin(\omega - \theta_{ij}) - S_j\sin\omega) \\ \times\; & (R\sin\theta_{jk} + S_k\sin(\omega + \theta_{jk}) - S_j\sin\omega) \\ +\; & (R\cos\theta_{ij} + S_j\cos(\omega - \theta_{ij}) - R - S_j\cos\omega) \\ \times\; & (R\cos\theta_{jk} + S_k\cos(\omega + \theta_{jk}) - R - S_j\cos\omega) \end{aligned}$$

This equation can be transformed into the following:

$$0 \;=\; (1 - \cos\theta_{ij} - \cos\theta_{jk} + \cos(\theta_{ij} + \theta_{jk}))R^2$$

$$
\begin{aligned}
&+ \ (2S_j - (S_j + S_k)\cos\theta_{ij} - (S_i + S_j)\cos\theta_{jk} \\
&\quad +(S_i + S_k)\cos(\theta_{ij} + \theta_{jk}))R\cos\omega \\
&+ \ ((S_k - S_j)\sin\theta_{ij} + (S_j - S_i)\sin\theta_{jk} \\
&\quad +(S_i - S_k)\sin(\theta_{ij} + \theta_{jk}))R\sin\omega \\
&+ \ S_j^2 + S_iS_k\cos(\theta_{ij} + \theta_{jk}) \\
&\quad -S_iS_j\cos\theta_{ij} - S_jS_k\cos\theta_{jk} \qquad (10)
\end{aligned}
$$

The values of $S_i$, $S_j$, $S_k$, $\theta_{ij}$, and $\theta_{jk}$ in Equation (10) are known. The equation can be expressed by the following linear form

$$K_1X_1 + K_2X_2 + K_3X_3 + K_4 = 0$$

where $K_i$, $i = 1, 2, 3, 4$, are the following coefficients:

$$
\begin{aligned}
K_1 &= 1 - \cos\theta_{ij} - \cos\theta_{jk} + \cos(\theta_{ij} + \theta_{jk}) \\
K_2 &= 2S_j - (S_j + S_k)\cos\theta_{ij} - (S_i + S_j)\cos\theta_{jk} + (S_i + S_k)\cos(\theta_{ij} + \theta_{jk}) \\
K_3 &= (S_k - S_j)\sin\theta_{ij} + (S_jS_i)\sin\theta_{jk} + (S_i - S_k)\sin(\theta_{ij} + \theta_{jk}) \\
K_4 &= S_j^2 + S_iS_k\cos(\theta_{ij} + \theta_{jk}) - S_iS_j\cos\theta_{ij} - S_jS_k\cos\theta_{jk}
\end{aligned}
$$

Furthermore, $X_i$, $i = 1, 2, 3$, are three linearly independent variables, with

$$
\begin{aligned}
X_1 &= R^2 \\
X_2 &= R\cos\omega \\
X_3 &= R\sin\omega
\end{aligned}
$$

Note that this linear form is the same as for the first alternative (distance constraint) except that the coefficients are different. This allows us to reuse the minimization approach as defined by Equation (9) and the way of calculating $R$ and $\omega$ as specified for the first alternative.

### 3.5.4. EXPERIMENTAL RESULTS

We briefly report about experiments with the line camera WAAC [Wide-Angle Airborne Camera, DLR; see (Reulke and Scheele, 1998) and Figure 3.9]. The specifications of the used model are as follows:

There are three (monochromatic) sensor lines on the focal plate of the camera, defining angles $\omega = 335°$, $\omega = 0°$ (the *center line*), and $\omega = 25°$. (For airborne missions these lines can be identified as forward looking line, nadir or center line, and backward looking line.) Rotating a WAAC on a tripod or turntable would actually allow to capture three panoramic images during one rotation. In the following experiments only image data captured via the backward line have been used. Each sensor line (image line) has


*Figure 3.7.* Test panorama with eight highlighted and indexed pairs of lines.

5,184 CCD cells (pixels). The effective focal length of the camera is $21.7mm$ for the center line, and $23.94mm$ for backward or forward line. The size (edge length) of one of WAAC's CCD cells is $0.007mm$. Thus, the value of $f_\mu$ (for the backward line) is equal to $3,420$ pixels. The camera was mounted on a turntable which supported an extension arm allowing $R$ to be up to $1.0m$.

Figure 3.7 shows a panoramic image taken in 2001 in a seminar room of the Institute for Space Sensor Technology and Planetary Exploration, at the German Aerospace Center (DLR) in Berlin. The size of this seminar room is about $120m^2$. The image has a resolution of $5,184 \times 21,388$ pixels. Eight pairs of lines are highlighted in the figure and indexed. They are used for estimating $R$ and $\omega$ based on the distance constraint. The value of $R$ was set to be $10cm$. The principal angle $\omega$ is equal to $155°$ in the sense of the definition of our general panoramic camera model (and $25°$ in terms of the WAAC specification as given in (Reulke and Scheele, 1998)). The lengths of used line segments were measured, with an expected error of no more than 0.5% of the length readings. All used data of these eight pairs of lines (shown in Figure 3.7) are summarized in Table 3.3.

The calibration proceeds as described above: use the optimization method of sequential quadratic programming (Gill et al., 1981) for estimating $R$ and

| Index | $H_i = H_j$ (m) | $h_i$ (pixel) | $h_j$ (pixel) | $D_{ij}$ (m) | $d_{ij}$ (pixel) |
|---|---|---|---|---|---|
| 1 | 0.0690 | 91.2 | 133.8 | 1.4000 | 1003.1 |
| 2 | 0.6320 | 600.8 | 683.0 | 1.0000 | 447.3 |
| 3 | 0.5725 | 351.4 | 367.4 | 1.5500 | 490.5 |
| 4 | 1.0860 | 1269.0 | 1337.6 | 0.6000 | 360.9 |
| 5 | 0.2180 | 273.0 | 273.6 | 0.2870 | 180.1 |
| 6 | 0.0690 | 81.8 | 104.2 | 1.4000 | 910.5 |
| 7 | 0.5725 | 318.0 | 292.0 | 1.5500 | 398.2 |
| 8 | 1.3300 | 831.2 | 859.4 | 1.3400 | 422.5 |

*Table 3.3:* Data of the eight pairs of lines shown in Figure 3.7.

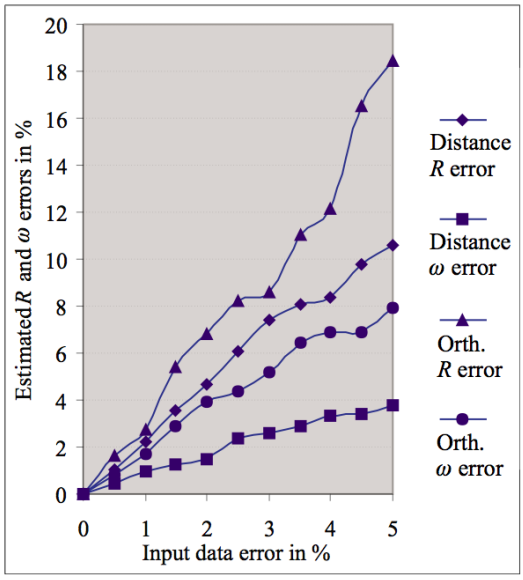



*Figure 3.8.* Error sensitivities for $R$ and $\omega$ when using either the distance constraint or the orthogonality constraint: the distance constraint performs better in both cases.

$\omega$, in particular minimize Equation (9). Results for the specified example are as follows: the use of all pairs of lines produces $R = 10.32cm$ and $\omega = 161.68°$, pairs {2,3,4,7,8} only then $R = 10.87cm$ and $\omega = 151.88°$, pairs {2,4,8} only then $R = 10.83cm$ and $\omega = 157.21°$. In general, such experiments showed that calibration accuracy is influenced by sample selection (e.g., aim at a uniform distribution of segments in the pictured scene) and quality of sample data (e.g., 0.5% error in length readings as in the given example is insufficient for high-accuracy calibration).

The error sensitivity of both alternatives of the parallel-line based approach can be evaluated by experiments using synthetic data. For example, we use (in an assumed scene geometry) one pair of lines for the first alternative, and one triple of orthogonal lines for the second. The simulated ground truth data are $R$= $10cm$ and $\omega = 155°$. We introduce errors to the ground-truth data $S_i$, $D_{ij}$, and $\theta_{ij}$, independently and with a maximum of 5% additive random noise in normal distribution. The range of $S_i$ is from $1m$ to $8m$, and the range of $\theta_{ij}$ is from 4° to 35°. The sample size is eight. Mean errors (for 100 trials in each case) are shown in Figure 3.8. Results suggest that estimated parameters using the orthogonality constraint are more sensitive to errors than in the case of using the distance constraint. The errors of the estimated parameters increase (only) linearly with respect to input errors for both cases. A combination of both constraints might be

useful.

Table 3.1 allows a performance comparison of the point-based, image correspondence, and parallel-line-based calibration approaches. The parallel-line-based approach performs best among these three (at their recent state of specification!). The parallel-line-based approach may also be based on a designed calibration object.

## References


Anandan, P., P. J. Burt, K. Dana, M. Hansen, and G. van der Wal: Real-time scene stabilization and mosaic construction. In Proc. *IEEE Works. Applications Computer Vision*, pages 54–62, 1994.

Arnspang, J., H. Nielsen, M. Chritensen, and K. Henriksen: Using mirror cameras for estimating depth. In Proc. *Computer Analysis Images Patterns*, LNCS 970, pages 711–716, 1995.

Avidan, S. and A. Shashua: Novel view synthesis in tensor space. In Proc. *IEEE Conf. Computer Vision Pattern Recognition* pages 1034–1040, 1997.

Ayache, N. J., and B. Faverjon: Efficient registration of stereo images by matching graph descriptions of edge segments. *Int. J. Computer Vision*, **1**:107–132, 1987.

Barnard, S. T., and M. A. Fischler: Computational stereo. *ACM Comput. Surv.*, **14**:553–572, 1982.

Baker. P., A. S. Ogale, and C. Fermüller. The Argus eye: a new imaging system designed to facilitate robotic tasks of motion. *IEEE Robotics Automation Magazine*, **11**:31–38, 2004.

Baker, S., and S. K. Nayar: A theory of single-viewpoint catadioptric image formation. *Int. J. Computer Vision*, **35**:1–22, 1999.

Bogner, S.: Introduction to panoramic imaging. In Proc. *IEEE SMC Conf.*, pages 3100–3106, 1995.

Burt, P. J., and E. H. Adelson: A multiresolution spline with applications to images mosaics. *ACM Trans. Graphics*, **2**:217–236, 1983.

Chahl, J. S., and M. V. Srinivassan. Reflective surfaces for panoramic visual sensor. *Applied Optics*, **36**:8275–8285, 1997.

Charles, J. R., R. Reeves, and C. Schur: How to build and use an all-sky camera. *Astronomy Magazine*, pages 64–70, April 1987.

Chen, C.-S., Y.-T. Chen, F. Huang: Stitching and reconstruction of linear-pushbroom panoramic images for planar scenes. In Proc. *European Conf. Computer Vision*, Volume 2, pages 190–201, 2004.

Chen, C.-Y., and R. Klette: Image stitching - comparisons and new techniques. In Proc. *Computer Analysis Images Patterns*, LNCS 1689, pages 615–622, 1999.

Chen, S. E.: QuickTimeVR - an image-based approach to virtual environment navigation. In Proc. *SIGGRAPH*, pages 29–38, 1995.

Cox, I. J.: A maximum likelihood n-camera stereo algorithm. In Proc. *Computer Vision Pattern Recognition*, pages 733–739, 1994.

Daniilides, K., and R. Klette (editors): *Computer Vision Beyond the Pinhole Camera*. Springer, Amsterdam, 2006.

Daniilides, K., and N. Papanikolopoulos (editors): *Panoramic Robots.* Special issue of *IEEE Robotics Automation Magazine*, Volume 11, December 2004.

Davis, J.E.: Mosaics of scenes with moving objects. In Proc. *Computer Vision Pattern Recognition*, pages 354–360, 1998.

Evers-Senne, J.-F., J. Woetzel, and R. Koch: Modelling and rendering of complex scenes with a multi-camera rig. In Proc. *European Conf. Visual Media Production*, pages 11–19, 2004.

Faugeras, O.: *Three-Dimensional Computer Vision: A Geometric Viewpoint.* The MIT Press, London, England, 1993.

Faugeras, O., and Q.-T. Luong. *The Geometry of Multiple Images.* The MIT Press, London, England, 2001.

Greene, N.: Environment mapping and other applications of world projections. *IEEE Trans. Computer Graphics Applications*, **6**:21–29, 1986.

Gimelfarb, G.: Probabilistic regularisation and symmetry in binocular dynamic programming stereo. *Pattern Recognition Letters*, **23**:431–442, 2002.

Gimelfarb, G. L., V. B. Marchenko, and V .I. Rybak: An algorithm for automatic identification of identical sections on stereo pairs of photographs. *Kibernetika*, no. 2, 118–129, 1972.

Gill, P. E., W. Murray, and M. H. Wright: *Practical Optimization.* Academic Press, London, 1981.

Gluckman, J., and S. K. Nayar. Ego-motion and omnidirectional cameras. In Proc. *Int. Conf. Computer Vision*, pages 999–1005, 1998.

Gluckman, J., and S. K. Nayar: Planar catadioptric stereo: Geometry and calibration. In Proc. *Computer Vision Pattern Recognition*, Volume I, pages 22–28, 1999.

Gluckman, J., S. K. Nayar, and K. J. Thorek: Real-time panoramic stereo. In Proc. DARPA, pages 299–303, 1998.

Goshtasby, A., and W. A. Grover: Design of a single-lens stereo camera system. *Pattern Recognition*, **26**:923–937, 1993.

Hartley, R. I.: Estimation of relative camera positions for uncalibrated cameras. In Proc. *European Conf. Computer Vision*, pages 579–587, 1992.

Hartley, R. and A. Zisserman: *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, United Kingdom, 2000.

Hecht, E., and A. Zajac: *Optics.* Addison-Wesley, New York, 1974.

Heyden, A.: Reconstruction from image sequences by means of relative depths. *Int. J. Computer Vision*, **24**:155–161, 1997.

Hilbert, D., and S. Cohn-Vossen: *Anschauliche Geometrie.* Springer-Verlag, Berlin, 1932.

Hong, J.: Image based homing. In Proc. *IEEE Int. Conf. Robotics Automation*, pages 620–625, 1991.

Huang, F., R. Klette, S. K. Wei, A. Börner, R. Reulke, M. Scheele, and K. Scheibe: Hyper-resolution and polycentric panorama acquisition and experimental data collection. Technical Report, CITR-TR-90, Dept. of Computer Science,The University of Auckland, 2001.

Huang, F., S. K. Wei, and R. Klette: Depth recovery system using object based layers. In Proc. *Image Vision Comp. New Zealand*, pages 199–204, 1999.

Huang, F., S. K. Wei, and R. Klette: Geometric analysis of symmetric panorama imaging. In Proc. *Image Vision Comp. New Zealand*, pages 19–24, 2001a.

Huang, F., S. K. Wei, and R. Klette: Geometrical fundamentals of polycentric panoramas. In Proc. *Int. Conf. Computer Vision*, Volume I, pages 560–565, 2001b.

Huang, F., S. K. Wei, and R. Klette: Stereo reconstruction from polycentric panoramas. In Proc. *Robot Vision*, LNCS 1998, pages 209–218, 2001c.

Huang, H.-C., and Y.-P. Hung: Panoramic stereo imaging system with automatic disparity warping and seaming. *Graph. Models Image Processing*, **60**:196–208, 1998.

Irani, M., P. Anandan, and S. Hsu: Mosaic based representations of video sequences and their applications. In Proc. *Int. Conf. Computer Vision*, pages 605–611, 1995.

Irani, M., S. Hsu, and P. Anandan: Video compression using mosaic representations. *Signal Processing: Image Communication*, **7**, pages 529–552, 1995a.

Ishiguro, H., T. Sogo, and T. Ishida: Human behavior recognition by a distributed vision system. In Proc. *DiCoMo Workshop*, pages 615–620, 1997.

Ishiguro, H., M. Yamamoto, and S. Tsuji: Omni-directional stereo. *IEEE Trans. Pattern Analysis Machine Intelligence*, **14**:257–262, 1992.

Jähne, B.: *Digital Image Processing: Concepts, Algorithms, and Scientific Applications,* 2nd edition. Springer, Berlin, 1993.

Kanatani, K.: *Geometric Computation for Machine Vision*. Oxford Univ. Press, New York, 1993.

Kang, S.-B., and P. K. Desikan: Virtual navigation of complex scenes using clusters of cylindrical panoramic images. Technical Report, CRL 97/5, Digital Equipment Corporation, Cambridge Research Lab, 1997.

Kang, S.-B., and R. Szeliski: 3-d scene data recovery using omnidirectional multibaseline stereo. *Int. J. Computer Vision*, **25**:167–183, 1997.

Kang, S.-B., and R. Weiss: Characterization of errors in compositing panoramic images. In Proc. *Computer Vision Pattern Recognition*, pages 103–109, 1997.

Klette, R., G. Gimel'farb, and R. Reulke. Wide-angle image acquisition, analysis and visualization. Invited plenary talk. In Proc. *Vision Interface*, pages 114-125, 2001.

Klette, R., and R. Reulke: Modeling 3D scenes: paradigm shifts in photogrammetry, remote sensing and computer vision. Keynote, *IEEE Int. Conf. Systems and Signals*, 8 pages, conference CD, 2005

Klette, R., and A. Rosenfeld. *Digital Geometry. Geometric Methods for Digital Picture Analysis*. Morgan Kaufmann, San Francisco, 2004.

Klette, R., and K. Scheibe: Combinations of range data and panoramic images - new opportunities in 3D scene modeling. Keynote, in Proc. *Computer Graphics Imaging Vision: New Trends*, pages 3 - 10, 2005.

Klette, R., K. Schlüns, and A. Koschan. *Computer Vision: Three-Dimensional Data from Images*. Springer, Singapore, 1998.

Kumar, R., P. Anandan, and K. Hanna: Direct recovery of shape from multiple views: A parallax based approach. In Proc. *Int. Conf. Pattern Recognition*, pages 685–688, 1994.

Kumar, R., P. Anandan, M. Irani, J. Bergen, and K. Hanna: Representation of scenes from collections of images. In Proc. *IEEE Works. Representation Visual Scenes*, pages 10–17, 1995.

Laveau, S., and O. Faugeras: 3-d scene representation as a collection of images. In Proc. *Int. Conf. Pattern Recognition*, pages 689–691, 1994.

Lee, M. C., W. G. Chen, C. L. B. Lin, C. Gu, T. Markoc, S. I. Zabinsky, and R Szeliski: A layered video object coding system using sprite and affine motion model. *IEEE Trans. Circuits Systems Video Techn.*, **7**:130–145, 1997.

Liu, G., R. Klette, and B. Rosenhahn: Structure from motion in the presence of noise. In Proc. *Image Vision Computing New Zealand* pages 138–143, 2005.

Luong, Q.-T., and T. Vieville: Canonic representations for the geometries of multiple projective views. In Proc. *European Conf. Computer Vision*, pages 589–599, 1994.

Mann, S., and R. W. Picard: Virtual bellows: Constructing high quality stills from video. In Proc. *IEEE Conf. Image Processing*, Volume I, pages 363–367, 1994.

Marr, D., and T. A. Poggio: A computational theory of human stereo vision. *Proc. R. Soc. Land. B.*, **204**:301–328, 1979.

McMillan, L., and G. Bishop: Head-tracked stereo display using image warping. In Proc. *Stereoscopic Displays Virtual Reality Systems*, SPIE-2409, pages 21–30, 1995.

McMillan, L., and G. Bishop: Plenoptic modeling: an image-based rendering system. In Proc. *SIGGRAPH*, pages 39–46, 1995a.

Murray, D.W.: Recovering range using virtual multicamera stereo. *Computer Vision Image Understanding*, **61**:285–291, 1995.

Nalwa, V.S.: A true omnidirectional viewer. Technical report, Bell Laboratories, Holmdel, NJ, February 1996.

Nayar, S.: Catadioptic omnidirectional cameras. In Proc. *Computer Vision Pattern Recognition*, pages 482–488, 1997.

Nayar, S.K.: Sphereo: Recovering depth using a single camera and two specular spheres. In Proc. *Optics Illumination Image Sensing Machine Vision*, SPIE, pages 251–263, 1998.

Nayar, S. K., and A. Karmarkar: 360 × 360 mosaics. In Proc. *Computer Vision Pattern Recognition*, Volume II, pages 388–395, 2000.

Nayar, S. K., and V. Peri: Folded catadioptric cameras. In Proc. *Computer Vision Pattern Recognition*, Volume II, pages 217–223, 1999.

Nene, S. A., and S. K. Nayar: Stereo with mirrors. In Proc. *Int. Conf. Computer Vision*, pages 1087–1094, 1998.

Neumann, J. and C. Fermüller. Plenoptic video geometry. *Visual Computer*, **19–26**:395–404, 2003.

Neumann, J., C. Fermüller, and Y. Aloimonos. Eyes from eyes: new cameras for structure from motion. In Proc. *Workshop on Omnidirectional Vision*, pages 19–26, 2002.

Neumann, J., C. Fermüller, and Y. Aloimonos. A hierarchy of cameras for 3d photography. *Computer Vision Image Understanding*, **96**:274–293, 2004.

Nishimura, T., T. Mukai, and R. Oka: Spotting recognition of gestures performed by people from a single time-varying image. In Proc. *Int. Conf. Robots Systems*, pages 967–972, 1997.

Ollis, M., H. Herman, and S. Singh: Analysis and design of panoramic stereo vision using equi-angular pixel cameras. Technical Report CMU-RI-TR-99-04, The Robotics Institute, Carnegie Mellon University, Pittsburgh, 1999.

Ohta, Y., and T. Kanade: Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Trans. Pattern Analysis machine Intelligence*, **7**:139–154, 1985.

Peleg, S., and M. Ben-Ezra: Stereo panorama with a single camera. In Proc. *Computer Vision Pattern Recognition*, pages 395–401, 1999.

Peleg, S., and J. Herman: Panoramic mosaic by manifold projection. In Proc. *Computer Vision Pattern Recognition*, pages 338–343, 1997.

Peleg, S., and J. Herman: Panoramic mosaics with videobrush. In Proc. *DARPA*, pages 261–264, 1997a.

Peleg, S., Y. Pritch, and M. Ben-Ezra: Cameras for stereo panoramic imaging. In Proc. *Computer Vision Pattern Recognition*, pages 208–214, 2000.

Petty, R., M. Robinson, and J. Evans: 3d measurement using rotating line-scan sensors. *Measurement Science Technology*, **9**:339–346, 1998.

Pollefeys, M., R. Koch, and L. Van Gool: Self?calibration and metric reconstruction in spite of varying and unknown intrinsic camera parameters. Int. J. Computer Vision, **32**:7–25, 1999.

Rademacher, R., and G. Bishop: Multiple-center-of-projection images. In Proc. *SIGGRAPH*, pages 199–206, 1998.

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

Sandau, R., and A. Eckardt: The stereo camera family waoss/waac for spaceborne/ airborne applications. *Int. Archives Photogrammetry Remote Sensing*, **XXXI(B1)**:170–175, 1996.

Šára, R.: The class of stable matchings for computational stereo. Technical Report CTU-CMP-1999-22, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, 1999.

Sawhney, H.S.: Simplifying motion and structure analysis using planar parallax and image warping. In Proc. *Int. Conf. Pattern Recognition*, Volume A, pages 403–408, 1994.

Sawhney, H. S., and S. Ayer: Compact representations of videos through dominant and multiple motion estimation. *IEEE Trans. Pattern Analysis machine Intelligence*, **18**:814–830, 1996.

Scheibe, K., H. Korsitzky, and R. Reulke: Eyescan - a high resolution digital panoramic camera. In Proc. *Robot Vision*, pages 77–83, 2001.

Seitz, S.M.: The space of all stereo images. In Proc. *Int. Conf. Computer Vision*, pages 26–33, 2001.

Shashua, A., and S. Avidan: The rank 4 constraint in multiple (over 3) view geometry. In Proc. *European Conf. Computer Vision*, Volume II, pages 196–206, April 1996.

Shashua, A., and M. Werman: Trilinearity of three perspective views and its associated tensor. In Proc. *Int. Conf. Computer Vision*, pages 920–925, 1995.

Shum, H.-Y., M. Han, and R. Szeliski: Interactive construction of 3d models from panoramic mosaics. In Proc. *Computer Vision Pattern Recognition*, pages 427–433, 1998.

Shum, H.-Y., and L.-W. He: Rendering with concentric mosaics. In Proc. *SIGGRAPH*, pages 299–306, 1999.

Shum, H., A. Kalai, and S. Seitz: Omnivergent stereo. In Proc. *Int. Conf. Computer Vision*, pages 22–29, 1999.

Shum, H.-Y., and R. Szeliski: Stereo reconstruction from multiperspective panoramas. In Proc. *Int. Conf. Computer Vision*, pages 14–21, 1999.

Szeliski, R., and S. Kang: Direct methods for visual scene reconstruction. In Proc. *IEEE Works. Representation Visual Scenes*, pages 26–33, 1995.

Southwell, D., J. Reyda, M. Fiala, and A. Basu: Panoramic stereo. In Proc. *Int. Conf. Pattern Recognition*, Volume A, pages 378–382, 1996.

Svoboda, T.: Central Panoramic Cameras Design, Geometry, Egomotion. PhD thesis, Czech Technical University, Prague, Czech Republic, 1999.

Svoboda, T., T. Pajdla, and V. Hlaváč: Epipolar geometry for panoramic cameras. In Proc. *European Conf. Computer Vision*, pages 218–232, 1998.

Szeliski, R.: Image mosaicing for tele-reality applications. In Proc. *IEEE Works. Applications Computer Vision*, pages 230–236, 1994.

Szeliski, R.: Video mosaics for virtual environments. *IEEE Trans. Computer Graphics Applications*, **16**:22–30, 1996.

Szeliski, R., and H.-Y. Shum: Creating full view panoramic image mosaics and environment maps. In Proc. *SIGGRAPH*, pages 251–258, 1997.

Thomas, G.: Real-time panospheric image dewarping and presentation for remote mobile robot control. *J. Advanced Robotics*, **17**:359–368, 2003.

Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. Robotics Automation*, **3**:323–344, 1987.

Valkenburg, R. J., X. Lin, and R. Klette: Self-calibration of target positions using geometric algebra. In Proc. *Image Vision Computing New Zealand*, pages 221-226, 2004.

Wei, S. K., F. Huang, and R. Klette: 3-d environment visualization by panoramic color anaglyphs. In Proc. *Image Vision Comp. New Zealand*, pages 280–285, 1998.

Wei, S. K., F. Huang, and R. Klette: Color anaglyphs for panorama visualizations. Technical Report, CITR-TR-19, Dept. of Computer Science, The University of Auckland, Auckland, 1998a.

Wei, S. K., F. Huang, and R. Klette: Three-dimensional scene navigation through anaglyphic panorama visualization. In Proc. *Computer Analysis Images Patterns*, LNCS 1689, pages 542–549, 1999.

Wei, S. K., F. Huang, and R. Klette: Three dimensional view synthesis from multiple images. Technical Report, CITR-TR-42, Dept. of Computer Science, The University of Auckland, Auckland, 1999a.

Wei, S. K., F. Huang, and R. Klette: Visibility determination in depth layered cylinder scene representations. In Proc. *Image Vision Comp. New Zealand*, pages 175–180, 1999b.

Wei, S. K., F. Huang, and R. Klette: The design of a stereo panorama camera for scenes of dynamic range. In Proc. *Int. Conf. Pattern Recognition*, Volume III, pages 635–638, 2002.

Wei, S. K., F. Huang, and R. Klette: Determination of geometric parameters for stereoscopic panorama cameras. *Machine Graphics & Vision*, **10**:399–427, 2002a.

Wei, S. K., M. Urban, and T. Pajdla: Stereo matching of catadioptric panoramic images. Technical Report CTU–CMP–2000–08, Center for Machine Perception, Czech Technical University, Prague, 2000.

Xu, G., and Z. Zhang: *Epipolar Geometry in Stereo, Motion and Object Recognition - A Unified Approach.* Kluwer, Amsterdam, 1996.

Yagi, Y.: Omnidirectional sensing and its applications. *IEICE Trans. Information Systems*, **E82-D**:568–579, 1999.

Yagi, Y.,and S. Kawato: Panoramic scene analysis with conic projection. In Proc. *IEEE/RSJ Int. Conf. Intelligent Robots Systems*, pages 181–187, 1990.

Yagi, Y., Y. Nishizawa, and M. Yachida: Map-based navigation for a mobile robot with omnidirectional image sensor copis. *IEEE Trans. Robotics Automation*, **11**:634–648, 1995.

Yamazawa, K., Y. Yagi, and M. Yachida: Omindirectional imaging with hyperboliodal projection. In Proc. *IEEE/RSJ Int. Conf. Intelligent Robots Systems*, pages 1029–1034, 1993.

Yamazawa, K., Y. Yagi, and M. Yachida: Obstacle detection with omnidirectional image sensor hyperomni vision. In Proc. *IEEE Int. Conf. Robotics Automation*, pages 1062–1067, 1995.

Zhang, Z., R. Deriche, O. Faugeras, and Q.-T. Luong: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, Lucioles, France, 1994.

Zhang, Z., and G. Xu: A general expression of the fundamental matrix for both perspective and affine cameras. In Proc. *Int. Joint Conf. Artificial Intelligence*, pages 23–29, Nagoya, Japan, August 1997.

Zheng, J.-Y., and S. Tsuji: Panoramic representation for route recognition by a mobile robot. *Int. J. Computer Vision*, **9**:55–76, 1992.

Zhu, Z., D. R. Karuppiah, E. M. Riseman, and A. R. Hanson: Keeping smart, omni-

directional eyes on you [adaptive panoramic stereovision]. *IEEE Robotics Automation Magazine*, **11**:69–78, 2004.