Foundations and Trends® in
Computer Graphics and Vision
Vol. 6, Nos. 1–2 (2010) 1–183

DOI: 10.1561/0600000023

Camera Models and Fundamental Concepts Used in Geometric Computer Vision

By Peter Sturm, Srikumar Ramalingam,
Jean-Philippe Tardif, Simone Gasparini
and João Barreto

Contents

Foundations and Trends® in
Computer Graphics and Vision
Vol. 6, Nos. 1–2 (2010) 1–183

DOI: 10.1561/0600000023

Camera Models and Fundamental Concepts Used in Geometric Computer Vision

Peter Sturm[1], Srikumar Ramalingam[2], Jean-Philippe Tardif[3], Simone Gasparini[4], and João Barreto[5]

[1] *INRIA Grenoble — Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, Montbonnot, France, Peter.Sturm@inrialpes.fr*
[2] *MERL, Cambridge, MA, USA, ramalingam@merl.com*
[3] *NREC — Carnegie Mellon University, Pittsburgh, PA, USA, tardifj@gmail.com*
[4] *INRIA Grenoble — Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, Montbonnot, France, Simone.Gasparini@inrialpes.fr*
[5] *Coimbra University, Coimbra, Portugal, jpbar@deec.uc.pt*

Abstract

This survey is mainly motivated by the increased availability and use of panoramic image acquisition devices, in computer vision and various of its applications. Different technologies and different computational models thereof exist and algorithms and theoretical studies for geometric computer vision ("structure-from-motion") are often re-developed without highlighting common underlying principles. One of the goals of this survey is to give an overview of image acquisition methods used in computer vision and especially, of the vast number of camera models that have been proposed and investigated over the years,

where we try to point out similarities between different models. Results on epipolar and multi-view geometry for different camera models are reviewed as well as various calibration and self-calibration approaches, with an emphasis on non-perspective cameras. We finally describe what we consider are fundamental building blocks for geometric computer vision or structure-from-motion: epipolar geometry, pose and motion estimation, 3D scene modeling, and bundle adjustment. The main goal here is to highlight the main principles of these, which are independent of specific camera models.

1

Introduction and Background Material

1.1 Introduction

Many different image acquisition technologies have been investigated in computer vision and other areas, many of them aiming at providing a wide field of view. The main technologies consist of catadioptric and fisheye cameras as well as acquisition systems with moving parts, e.g., moving cameras or optical elements. In this monograph, we try to give an overview of the vast literature on these technologies and on computational models for cameras. Whenever possible, we try to point out links between different models. Simply put, a computational model for a camera, at least for its geometric part, tells how to project 3D entities (points, lines, etc.) onto the image, and vice versa, how to back-project from the image to 3D. Camera models may be classified according to different criteria, for example the assumption or not of a single viewpoint or their algebraic nature and complexity. Also, recently several approaches for calibrating and using "non-parametric" camera models have been proposed by various researchers, as opposed to classical, parametric models.

In this survey, we propose a different nomenclature as our main criterion for grouping camera models. The main reason is that even

so-called non-parametrics models do have parameters, e.g., the coordinates of camera rays. We thus prefer to speak of three categories: (i) A **global camera model** is defined by a set of parameters such that changing the value of any parameter affects the projection function all across the field of view. This is the case for example with the classical pinhole model and with most models proposed for fish-eye or catadioptric cameras. (ii) A **local camera model** is defined by a set of parameters, each of which influences the projection function only over a subset of the field of view. A hypothetical example, just for illustration, would be a model that is "piecewise-pinhole", defined over a tessellation of the image area or the field of view. Other examples are described in this monograph. (iii) A **discrete camera model** has sets of parameters for individual image points or pixels. To work with such a model, one usually needs some interpolation scheme since such parameter sets can only be considered for finitely many image points. Strictly speaking, discrete models plus an interpolation scheme are thus not different from the above local camera models, since model parameters effectively influence the projection function over regions as opposed to individual points. We nevertheless preserve the distinction between discrete and local models, since in the case of discrete models, the considered regions are extremely small and since the underlying philosophies are somewhat different for the two classes of models.

These three types of models are illustrated in Figure 1.1, where the camera is shown as a black box. As discussed in more detail later in the monograph, we mainly use back-projection to model cameras, i.e., the mapping from image points to camera rays. Figure 1.1 illustrates back-projection for global, discrete and local camera models.

After describing camera models, we review central concepts of geometric computer vision, including camera calibration, epipolar and multi-view geometry, and structure-from-motion tasks, such as pose and motion estimation. These concepts are exhaustively described for perspective cameras in recent textbooks [137, 213, 328, 336, 513]; our emphasis will thus be on non-perspective cameras. We try to describe the various different approaches that have been developed for camera calibration, including calibration using grids or from images of higher level primitives, like lines and spheres, and self-calibration. Throughout

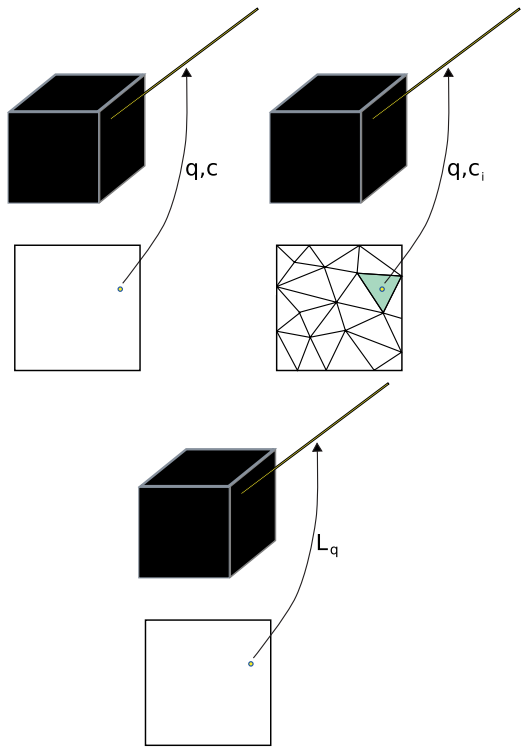


Fig. 1.1 Types of camera models. *Left*: For global models, the camera ray associated with an image point $\mathbf{q}$ is determined by the position of $\mathbf{q}$ and a set of global camera parameters contained in a vector $\mathbf{c}$. *Middle*: For discrete models, different image regions are endowed with different parameter sets. *Right*: For discrete models, the camera rays are directly given for sampled image points, e.g., by a look-up table containing Plücker coordinates, here the Plücker coordinates $\mathbf{L_q}$ of the ray associated with image point $\mathbf{q}$.

this monograph, we aim at describing concepts and ideas rather than all details, which may be found in the original references.

The monograph is structured as follows. In the following section, we give some background material that aims at making the mathematical treatment presented in this monograph, self-contained. In Section 2, we review image acquisition technologies, with an emphasis on omnidirectional systems. Section 3 gives a survey of computational camera models in the computer vision and photogrammetry literature, again emphasizing omnidirectional cameras. Results on epipolar and multi-view geometry for non-perspective cameras are summarized

in Section 4. Calibration approaches are explained in Section 5, followed by an overview of some fundamental modules for structure-from-motion in Section 6. The monograph ends with conclusions, in Section 7.

1.2 Background Material

Given the large scope of this monograph, we rather propose summaries of concepts and results than detailed descriptions, which would require an entire book. This allows us to keep the mathematical level at a minimum. In the following, we explain the few notations we use in this monograph. We assume that the reader is familiar with basic notions of projective geometry, such as homogeneous coordinates, homographies, etc. and of multi-view geometry for perspective cameras, such as the fundamental and essential matrices and projection matrices. Good overviews of these concepts are given in [137, 213, 328, 336, 513].

Fonts. We denote scalars by italics, e.g., s, vectors by bold characters, e.g., $\mathbf{t}$ and matrices in sans serif, e.g., A. Unless otherwise stated, we use homogeneous coordinates for points and other geometric entities. Equality between vectors and matrices, up to a scalar factor, is denoted by $\sim$. The cross-product of two 3-vectors $\mathbf{a}$ and $\mathbf{b}$ is written as $\mathbf{a} \times \mathbf{b}$.

Plücker coordinates for 3D lines. Three-dimensional lines are represented either by two distinct 3D points, or by 6-vectors of so-called Plücker coordinates. We use the following convention. Let $\mathbf{A}$ and $\mathbf{B}$ be two 3D points, in homogeneous coordinates. The Plücker coordinates of the line spanned by them, are then given as:

$$\begin{pmatrix} B_4\bar{\mathbf{A}} - A_4\bar{\mathbf{B}} \\ \bar{\mathbf{A}} \times \bar{\mathbf{B}} \end{pmatrix}, \tag{1.1}$$

where $\bar{\mathbf{A}}$ is the 3-vector consisting of the first three coordinates of $\mathbf{A}$ and likewise for $\bar{\mathbf{B}}$.

The action of displacements on Plücker coordinates is as follows. Let $\mathbf{t}$ and R be a translation vector and rotation matrix that map points according to:

$$\mathbf{Q} \mapsto \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{Q}.$$

Plücker coordinates are then mapped according to:

$$\mathbf{L} \mapsto \begin{pmatrix} \mathsf{R} & \mathsf{0} \\ -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \end{pmatrix} \mathbf{L}, \tag{1.2}$$

where $\mathsf{0}$ is the 3×3 matrix composed of zeroes.

Two lines cut one another exactly if

$$\mathbf{L}_2^\mathsf{T} \begin{pmatrix} \mathsf{0}_{3\times3} & \mathsf{I}_{3\times3} \\ \mathsf{I}_{3\times3} & \mathsf{0}_{3\times3} \end{pmatrix} \mathbf{L}_1 = 0. \tag{1.3}$$

Lifted coordinates. It is common practice to linearize polynomial expressions by applying Veronese embeddings. We use the informal term "lifting" for this, for its shortness. Concretely, we apply lifting to coordinate vectors of points. We will call "n-order lifting" of a vector $\mathbf{a}$, the vector $\mathcal{L}^n(\mathbf{a})$ containing all n-degree monomials of the coefficients of $\mathbf{a}$. For example, second and third order liftings for homogeneous coordinates of 2D points, are as follows:

$$\mathcal{L}^2(\mathbf{q}) \sim \begin{pmatrix} q_1^2 \\ q_1 q_2 \\ q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix} \quad \mathcal{L}^3(\mathbf{q}) \sim \begin{pmatrix} q_1^3 \\ q_1^2 q_2 \\ q_1 q_2^2 \\ q_2^3 \\ q_1^2 q_3 \\ q_1 q_2 q_3 \\ q_2^2 q_3 \\ q_1 q_3^2 \\ q_2 q_3^2 \\ q_3^3 \end{pmatrix}. \tag{1.4}$$

Such lifting operations are useful to describe several camera models. Some camera models use "compacted" versions of lifted image point coordinates, for example:

$$\begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}.$$

We will denote these as $\bar{\mathcal{L}}^2(\mathbf{q})$, and use the same notation for other lifting orders.

2

Technologies

We briefly describe various image acquisition technologies, with an emphasis on omnidirectional ones. This section aims at describing the most commonly used technologies in computer vision and related areas, without any claim of being exhaustive. More information, including historical overviews, can be found in the following references, which include textbooks, articles, and webpages [50, 51, 54, 71, 103, 220, 227, 294, 326, 335, 528, 541].

2.1 Moving Cameras or Optical Elements

2.1.1 Slit Imaging

Slit imaging has been one of the first techniques to acquire panoramic images. Various prototypes existed already in the nineteenth century, usually based on a moving slit-shaped aperture. Historical overviews are given in the references in the first paragraph of this section. In the following, we only review some more recent, digital slit imaging systems, mainly those developed for robotic and computer vision; similar systems were also developed for photogrammetric applications [326].

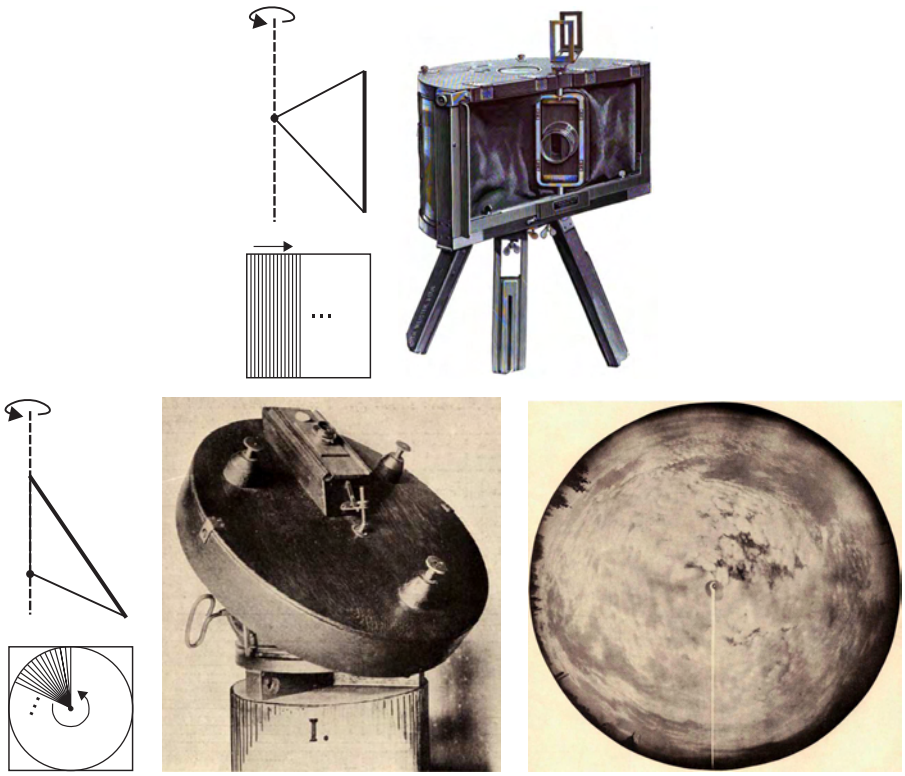


Fig. 2.1 Examples of generation of central slit images. *Top*: "Standard" slit imaging principle and an early realization of it, the cylindrograph of Moëssard [351]. *Bottom*: Slit imaging with a tilted 1D sensor, the so-called "cloud camera" and an image acquired with it (see text).

Most of these systems either use a 2D camera or a 1D camera (also called linear camera or pushbroom camera) which "scans" a scene while moving, generating a panoramic image (cf. Figure 2.1). In the 2D camera case, only one or several columns or rows of pixels are usually kept per acquired image, and stitched together to form a panoramic image. Note that pushbroom images are highly related to the so-called epipolar plane images, see for example [56].

Sarachik, Ishiguro et al., and Petty et al. acquired panoramas from a rotating perspective camera by glueing together pixel columns from each image, and used them for map building or 3D measurement [247, 400, 434]. Barth and Barrows developed a panoramic image acquisition

system based on a fast panning linear camera, thus significantly decreasing acquisition times [38]. A similar system was developed by Godber et al. [178]. Benosman et al. developed a panoramic stereo sensor based on the same principle, consisting of two linear cameras mounted vertically on top of each other [52]; they rotate together about their baseline and panoramic images are generated by stacking the acquired 1D images together. Issues of calibration, matching, and 3D reconstruction are addressed in [52]. Klette et al. reviewed several design principles and applications of such devices [281].

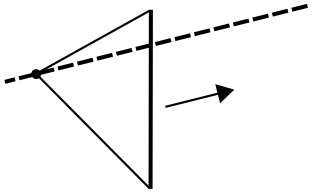
Usually, the rotation axis is parallel to the "slits" of pixels used to form the panoramic image. A historic example of rotating a tilted camera, is a design described by Fassig [136], called "cloud camera": it acquires a hemispheric field of view by revolving a tilted camera with a wedge-shaped opening around an axis (cf. Figure 2.1).

There are many more such systems; an exhaustive list is out of scope. The above systems are all designed to deliver **central panoramic images**: panoramic referring to omnidirectionality in one orientation, central referring to images having a single effective point of view. This is achieved by rotating the 1D or 2D camera about an axis containing its optical center.

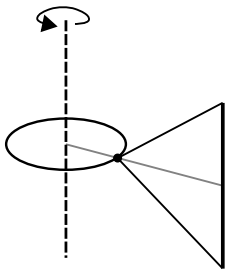
In the following, we review several **non-central slit imaging systems**.

One-dimensional cameras, or pushbroom cameras, are routinely used in satellite imaging, since they provide a cheap way of obtaining high-resolution images and since they are well adapted to the way satellites scan planets. Pushbroom panoramas are non-central since each column is acquired at a different position of the satellite. The special case of linear pushbroom panoramas, where the camera motion is assumed to be a straight line (cf. Figure 2.2(a)), was extensively modeled by Gupta and Hartley [197], see Section 3.1.4. They also proposed a sensor model and a calibration method for the case of pushbroom cameras on an orbiting satellite, e.g., moving along an elliptical trajectory [198].

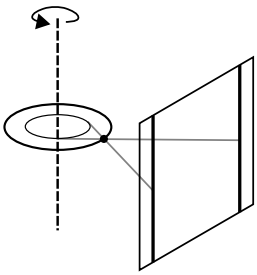
Concerning close-range applications, Zheng and Tsuji seem to be among the first researchers to have introduced and used non-central panoramic images [570, 571, 572], sometimes also called non-central mosaics, motion panoramas, or omnivergent images. Like in other



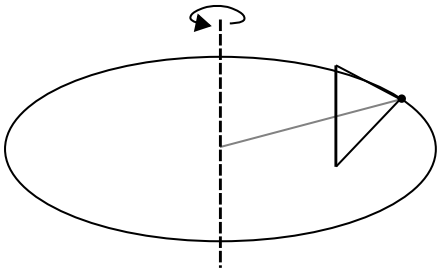
(a) Linear pushbroom acquisition.



(b) Circular non-central panorama.



(c) Dual slit principle to generate stereo panoramas using one rotating 2D camera.



(d) Cyclograph generation with inward looking camera.

(e) Example of a cyclograph, courtesy of Steve Seitz [444].

(f) Example of a panorama obtained with the approach of [4], see text. Courtesy of Aseem Agarwala.

Fig. 2.2 Several ways of generating non-central slit or slit-like images.

systems, they proceeded by acquiring images using a moving camera, through a vertical slit, and stacking the acquired slit images one next to the other. If the camera is rotating about the vertical axis through its optical center, then the acquired image is a cylindrical mosaic, or panorama. If the camera is rotating about some other vertical axis, then we obtain non-central mosaics (cf. Figure 2.2(b)). Zheng and Tsuji have also proposed a generalization of this principle, for a camera moving on any smooth path. In the case of a straight line, we of course find the

above linear pushbroom panoramas. Zheng and Tusji used such motion panoramas for route following of robots. Besides explaining the generation of panoramas, they also analyzed the apparent distortions in the obtained images and other issues. They also used dual-slit panoramas (two panoramas acquired using the same camera and two vertical slits) for stereo computations (cf. Figure 2.2(c)), an idea that was later generalized by Peleg et al. [394, 395, 397] and Li et al.[314], Seitz et al. [444]. Here, 3D points can be reconstructed from point matches between the two non-central images by triangulation.

Ishiguro et al. used such panoramic views for map generation [246, 247]. They also used the dual-slit panoramas explained above for stereo computations, as well as stereo from panoramas acquired at different positions.

McMillan and Bishop used panoramic images for image-based rendering, inspired by the plenoptic function concept [338].

Krishnan and Ahuja studied how to obtain the sharpest panoramic images from a panning camera [288]. They showed that when using a regular camera, whose CCD is perpendicular to the optical axis, the camera should be rotated about an off-center point on the optical axis, together with acquiring images with a varying focus setting. This effectively yields non-central panoramas. Krishnan and Ahuja also proposed another sensor design, where the CCD plane is not perpendicular to the optical axis, and used it for panoramic image acquisition. The advantage of such a sensor is that the depth of field volume is skewed and so, while panning with a *fixed* focus setting, the union of the images' depth of field volumes, is larger than for a fronto-parallel CCD, hence avoiding to vary the focus setting while panning.

Usually, cameras are facing outward when acquiring slit images. Inward facing cameras were also considered, as early as in 1895 [105], leading to the principle of peripheral photography or images called cyclographs. A sample cyclograph obtained using this approach is shown in Figure 2.2(e). Let us mention another acquisition principle that does not produce slit panoramas but is considering the acquisition of images taken around an object: Jones et al. proposed a setup to acquire images as if taken all around an object, with a static camera and object [258], whereas the usual approach is to use multiple cameras or a

turntable. To do so, they placed a cylindrical mirror around the object and used a rotating planar mirror to select the effective viewpoint.

Peleg et al. showed how to acquire non-central panoramas without having to move the camera, by using a mirror of a particular shape or a lens with a particular profile [396, 395]. When combining a static central camera with the appropriate mirror or lens, the camera rays of the compound system are distributed the same way as in circular non-central panoramas, i.e., they are incident with a circle containing the camera's optical center (cf. Figure 2.2(b)).

Seitz et al. studied these ideas with the goal of obtaining the most accurate 3D reconstruction from possibly non-central stereo pairs [444]. Their problem definition is that only a 2D set of lines of sight may be acquired as an image, from viewpoints constrained within a sphere. In order to maximize the reconstruction accuracy of points outside the sphere, each point should be viewed along two lines of sight at least, with maximum vergence angle. The result is that the lines of sight to be stored are tangents of the sphere. How to best choose the appropriate tangents is described in [444], resulting in so-called omnivergent images. Interestingly, omnivergent stereo pairs, although being pairs of non-central images, do have a standard epipolar geometry, with horizontal epipolar lines, thus allow to directly apply standard stereo algorithms (more on this in Section 3.4.2). However, acquiring such spherical omnivergent images is not simple, although Nayar and Karmarkar proposed systems for this task, based on catadioptric or fisheye cameras [371]. An approximate omnivergent stereo pair can be acquired via the dual-slit principle introduced by Zheng et al., see above.

Agarwala et al. have extended the principle of using slits of images to compose panoramic images, toward the use of image regions, chosen in order to better conform to the scene's shape [4]. They apply this to generate multi-viewpoint panoramas of urban scenes, such as of the facades of buildings along streets (see Figure 2.2(f) for an example).

Ichimura and Nayar studied the problem of motion and structure recovery from freely moving 1D sensors and rigid rigs of two or three 1D sensors, as well as from special motions; their study subsumes several previously studied cases such as linear pushbroom panoramas [243].

2.1.2 Classical Mosaics

By "classical" mosaics, we refer to classical in terms of the computer vision community, i.e., mosaics generated by stitching together 2D images, but noting that slit imaging to generate mosaics has been done before, at least with analog cameras. A tutorial on image alignment and stitching has been published in the same journal as this survey, by Szeliski [487]. Due to this good reference and the fact, that classical mosaic generation is widely known, we do not describe this any further and simply give a few additional references.

An early and often overlooked work on digital mosaics is by Yelick and Lippman, who showed how to combine images obtained by a rotating camera to generate a mosaic [322, 547]. In his bachelor thesis [547], Yelick also discussed other omnidirectional image acquisition techniques, such as the fisheye lens and rotating slit cameras. Later works on digital mosaics include those by Teodosio and Mills [496], Teodosio and Bender [495] and Chen [92], to name but a few among the many existing ones.

2.1.3 Other Technologies

Murray analyzed the setup by Ishiguro et al. and others (see Section 2.1.1) and proposed an alternative solution where instead of rotating the camera about some axis, the camera looks at a planar mirror which rotates about an axis [362]. Such a system was already proposed for aerial imaging by Bouwers and van der Sande [58]. However, the issue of how to compensate for the camera displacement during rotations of the mirror, due to the airplane's motion, was not fully discussed. Other systems that use rotating mirrors or prisms are referenced by Yagi [541], see also Section 2.3.4.

2.2 Fisheyes

The concept of fisheye view and lens dates back to more than a century [48, 226, 350, 536]. Fisheye lenses or converters can achieve larger than hemispheric fields of view but are usually still relatively costly. A technical report suggesting inexpensive simple solutions to build

fisheye lenses was provided by Dietz [115]. Some more references on fisheyes are given in Section 3.1.7.

2.3 Catadioptric Systems

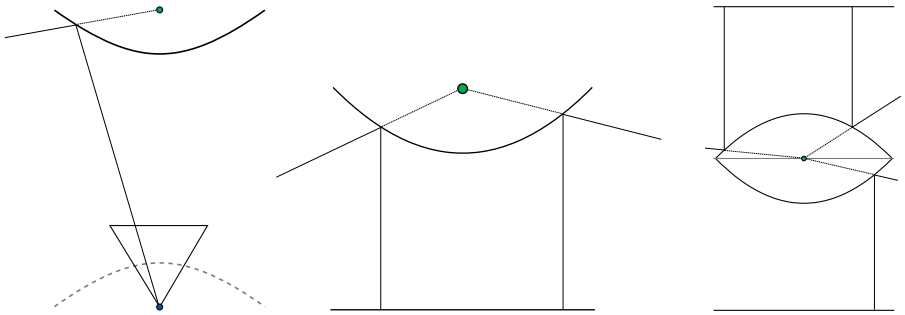
Using external mirrors together with cameras allows for a broad range of design possibilities, which is one of the reasons of the large number of catadioptric devices proposed alone in the computer vision community. Camera design may follow different goals, foremost in our context being a wide field of view, others being for example the compactness of a sensor, a single effective viewpoint, image quality, focusing properties, or a desired projection function. In the following, we exclusively describe geometric properties, in terms of projection function (focusing etc. of course also depend on mirror geometry).

We may distinguish five types of catadioptric systems: (i) single-mirror central systems, having a single effective viewpoint, (ii) central systems using multiple mirrors, (iii) non-central systems, (iv) single-lens stereo systems, and (v) programmable devices.

The fourth category, single-lens stereo systems, is of course a subset of the third category, non-central systems, but is singled out here since the goal is to obtain a sufficiently non-central system in order to enable accurate 3D modeling whereas for other systems, "non-centrality" is usually not a design goal but an artifact or a consequence of other design goals.

2.3.1 Single-Mirror Central Catadioptric Systems

Baker and Nayar derived all single-mirror central catadioptric systems [22] (Bruckstein and Richardson obtained some of these results independently [66]). Essentially the same results were also known outside the scope of catadioptric vision, e.g., were reported in 1637 by Descartes [112] and later by Feynman et al. [142] and Drucker and Locke [121]; further, it is likely that they were already known in antiquity to Greek geometers. Single-mirror central catadioptric systems can only be constructed using mirrors whose surface is obtained by revolving a conic about a symmetry axis of the conic. In addition, the camera looking at the mirror must be central and its optical center



(a) Hyper-catadioptric camera. The dashed curve is the second sheet of the hyperbola.

(b) Para-catadioptric camera.

(c) Two para-cata systems put back-to-back such that their effective viewpoints coincide.

Fig. 2.3 Illustration of (compositions of) single-mirror central catadioptric systems. Effective viewpoints are shown in green and the position of the true camera, in blue. The latter is only shown for the hyper-catadioptric case, for the para-catadioptric one, the camera is telecentric, thus has an optical center at infinity.

has to coincide with one of the conic's foci, otherwise the whole system is non-central.[1] The other focus is then the effective viewpoint of the catadioptric system: any (back-projection) line going out from the camera center goes, after being reflected in the mirror, through that second focus (cf. Figure 2.3).

The following special cases exist: hyperboloidal (cf. Figure 2.3(a)), paraboloidal (cf. Figure 2.3(b)), ellipsoidal, cone-shaped or planar mirrors. As for paraboloidal mirrors, one of the two real focus points is a point at infinity. In order to obtain a wide field of view, the only option is to "position" the camera at that point, which can be achieved using a telecentric lens. As for cone-shaped mirrors, the camera's optical center has to be located at the cone's tip; hence the only part of the mirror the camera sees corresponds to the rays that graze the mirror surface. Cone-shaped mirrors are thus theoretically excluded from the set of

[1] It may be possible to achieve a central catadioptric system using a *non-central* camera looking at a mirror which is not necessarily conic-based; all that matters is that the rays that are back-projected from the non-central camera converge to a single point after reflection in the mirror. For any non-central camera and a desired effective viewpoint, one may be able to design an appropriate mirror.

useful central catadioptric mirrors, although it has been shown that, with a more general modeling of optics than applied in [22], practical central viewing can still be achieved, see further below in this section. Ellipsoidal mirrors do not allow to increase the field of view and are thus not directly appropriate to build omnidirectional cameras; however, they are useful in designing so-called folded catadioptric systems, consisting of two or more mirrors (see Section 2.3.2). As for spherical mirrors, a special case of ellipsoidal ones, both real foci lie at the sphere center; a camera positioned there only sees the reflection of itself, which makes this case impractical. Overall, the only two systems deemed generally practical are the para-catadioptric and hyper-catadioptric ones, based on paraboloidal and hyperboloidal mirrors.

Central hyper-catadioptric systems seem to have been used first, cf. the patent by Rees in 1970 [424] and first applications in robotics and computer vision by Yagi and his co-workers [540, 546]. They highlighted the single viewpoint property if the camera is positioned at one of the mirror's foci and also discussed optical properties of the system such as blur. They also combined cameras of different types, such as in the MISS system which is composed of a cone-based catadioptric camera (non-central) and a standard stereo system [544]. This allows to combine mutual advantages of the sensors, such as good localization using the catadioptric camera and larger resolution with the stereo system. A full trinocular analysis of line images was also proposed.

Nayar introduced central para-catadioptric systems, consisting of an orthographic camera and a parabolic mirror, positioned such that the viewing direction of the camera is parallel to the mirror axis [367, 368]. He also showed that two para-catadioptric sensors with a field of view of 180° each, can be put back-to-back to achieve a full spherical field of view while still preserving a single effective optical center (cf. Figure 2.3(c)). This is possible since a 180° field of view is achieved with a para-catadioptric system if the mirror extends till the cross section containing its finite focus. Since the finite focus is exactly the effective optical center, putting two such mirrors back-to-back makes their finite foci coincide and thus also the two effective optical centers.

Such systems are rather widely used nowadays and commercialized by several companies, see e.g., [103].

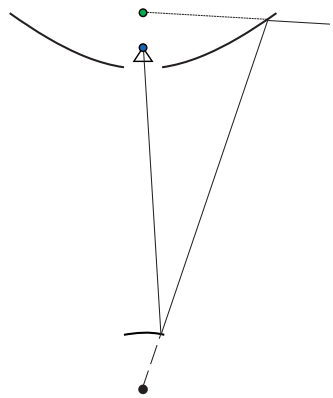
As mentioned above, cone-shaped mirrors were predicted as being impractical to achieve central catadioptric systems. This issue was reconsidered by Lin and Bajcsy [319, 320]. Their starting point was that the geometric analysis of Baker and Nayar considered that each 3D point is imaged along a single light path, whereas due to the finite aperture of real lenses, light rays emitted by a point within a finite volume, hit the image area. Reciprocally, a camera whose main lens is located at the tip of a cone (the actual tip being cut off) actually sees the mirror surface and not only rays grazing it. Lin and Bajcsy showed that it is possible to obtain sharp single-viewpoint catadioptric images in this case.

2.3.2 Central Catadioptric Cameras with Multiple Mirrors — Folded Catadioptric Cameras

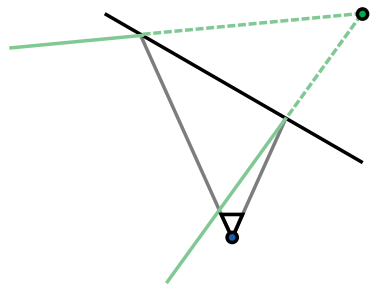
Central catadioptric cameras can also be achieved when using more than one mirror. This allows more compact sensor designs and gives additional degrees of freedom to improve optical properties. Such designs are also termed *folded catadioptric cameras*; an excellent reference is [372], where Nayar and Peri presented several designs, design issues, and references to other works in this area.

Roughly speaking, when combining conic-shaped mirrors, and positioning them such that foci of successive mirrors in a sequence of reflections, coincide, then central image acquisition is possible: by placing a camera at the left-over focus of the "last" mirror, the compound system has a single effective viewpoint at the left-over focus of the "first" mirror (cf. Figure 2.4(a)).

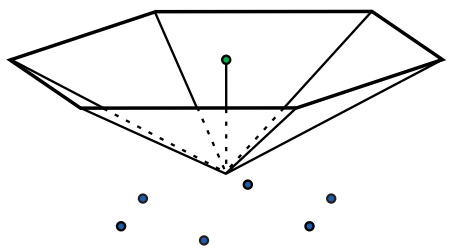
Previously, Yagi and Yachida proposed such a system, consisting of two paraboloidal mirrors [542]. Nagahara et al. proposed a design for catadioptric optics for head mounted devices (HMDs), consisting of three mirrors, a planar, a hyperboloidal, and an ellipsoidal one, arranged such as to provide a single effective viewpoint, while achieving a wide field of view and avoiding occlusions of the field of view by the mirrors themselves [363]. Takeya et al. proposed another similar design [488]. Kim and Cho addressed the problem of calibrating a system composed of multiple successive mirrors, using a learning-based



(a) A folded central catadioptric system based on two hyperboloidal mirrors. The camera is now located inside one of the mirrors, at the blue position. The effective viewpoint of the compound systems is shown in green.



(b) A camera looking at a planar mirror is identical to a virtual camera behind the mirror, looking "the other way round". Shown is the field of view of the true and the virtual camera; the green point is the effective viewpoint.



(c) Assembling camera–mirror pairs such that effective viewpoints coincide.

Fig. 2.4 Illustration of multi-mirror central catadioptric systems.

approach [280]. Nagahara et al. showed how to achieve uniform angular resolution for a folded catadioptric camera with two mirrors and a single effective viewpoint [364].

Another way of achieving central projection with multiple mirrors is what is often called the "Nalwa pyramid", although an earlier patent on essentially the same system is due to Iwerks [250]. Iwerks, Nalwa and independently, Kawanishi et al., proposed to use several regular cameras and as many planar mirrors [250, 271, 365]. A camera looking at a planar mirror produces the same image (up to side-swapping)

as a camera located behind the mirror, at the position obtained by reflecting the original center in the mirror, and "inversely" oriented (cf. Figure 2.4(b)). The reflected optical center position is the effective viewpoint here. The main idea consists in arranging camera–mirror pairs such that the effective viewpoints coincide. This is the easiest done in a pyramidal layout (cf. Figure 2.4(c)), but others are imaginable. Nalwa's original design consists of four camera–mirror pairs, Kawanishi et al. used six. Main advantages of such a system are omnidirectional viewing (the different camera images can be stitched together), a single effective viewpoint, and high resolution (as compared to mono-camera omnidirectional systems).

Gao et al. propose a similar design that has a hemispherical field of view [155]. Majumder et al. and Hua et al. placed two mirror–camera pyramids back-to-back such that the effective viewpoints of the two pyramids coincide, thus also enhancing the vertical field of view of the compound sensor [234, 235, 329].

Greguss developed the so-called panoramic annular lens (PAL) which combines reflective and refractive elements in a compact layout and achieves panoramic viewing with a full spherical field of view in horizontal direction and a vertical field of view of about 40° [188, 425]. Greguss' design has been used and/or improved for example by Powell [407] and Zhu et al. [574].

Yin and Boult built a sensor consisting of a tree of three different-sized coaxial paraboloidal mirrors and an orthographic camera looking at them [549]. Their motivation was not to achieve a single-lens stereo system, rather to obtain an image pyramid by optical means: the three mirrors are chosen such that they lead to omnidirectional images where each one doubles the resolution with respect to the previous one. Ideally, the mirrors' foci should coincide in order for the three pyramid images to correspond to the same viewpoint; this is not exactly the case in the proposed system but it was shown that the misalignment is negligible if the scene is sufficiently far away.

2.3.3 Non-central Catadioptric Cameras

Spherical mirrors. Hong et al. used a non-central catadioptric system with a spherical mirror for robot homing, a visual servoing

task [231]. They actually did not use the entire field of view that the camera offers, but just a region around the horizon plane; the horizon plane, i.e., the plane swept out by horizontal back-projection rays, corresponds to a circle in the image plane. The robot on which the camera is mounted is supposed to move on a horizontal ground plane. Hence, points in the horizon plane are always imaged on the above circle. In order to perform homing, the authors thus exploit a narrow region around that circle, perform image analysis and matching in it and use the result for a 2D visual servoing (translation and rotation in the ground plane). While a spherical mirror always leads to a non-central catadioptric system (unless the camera is placed at the sphere center), considering only the image portion corresponding to the horizon plane, actually corresponds to using an omnidirectional 1D camera with a single effective viewpoint.

Cone-shaped mirrors. Yagi and Kawato used a catadioptric system with a cone-shaped mirror (called COPIS) for robot motion estimation and map building [542]. The system was non-central since the camera was located at a certain distance from the tip of the cone. Yagi and Kawato described the forward projection model as well as 3D point triangulation and epipolar geometry for translational motion. They as well as other researchers have used cone-shaped mirrors mounted vertically on robots since vertical lines, which are omnipresent in man-made scenes, are imaged as (long) radial lines and are thus easy to extract, e.g., [61, 83, 321, 393, 456]. Mouaddib and his co-workers used this idea for pose estimation, motion estimation, map building, etc. with the sensor SYCLOP they developed [393]. They proposed a calibration method, where the sensor is immersed in a hollow cube with a calibration checkerboard painted on the inside of each face [61, 83]. They then use both, extracted points and vertical lines of the calibration object, for calibration. A method for pose estimation (or, absolute localization) is presented in [82].

Purpose-made mirrors optimizing resolution or satisfying other design goals. A drawback of "standard" catadioptric sensors is a significant variation of effective resolution, or spatial/angular resolution, across the image area. Several works aimed at removing

or alleviating this, see the nice overview by Hicks and Perline [225]. Different works aimed at achieving different properties for sensor resolution. Uniform angular resolution is achieved by what is usually called equidistant or equiangular projection: the angle spanned by the back-projection ray of an image point and the optical axis, is proportional to the distance between the image point and the principal point. This is approximately the case for most fisheye lenses. Ollis et al. showed how to compute mirror surfaces that provide an equiangular catadioptric sensor [386]. Previously, Chahl and Srinivasan obtained a similar result, a mirror shape where the angle between the optical axis and a ray back-projected from the camera, is proportional to the same angle after reflecting the ray in the mirror [84]. Conroy and Moore achieved a sensor with solid angle pixel density invariance, i.e., where the surface area of a circular image portion is proportional to the solid angle of the corresponding field of view [99]. A similar work is due to Hicks and Perline [225]. Another related work by Gächter et al. achieved a similar goal [153], but for the case where the camera looking at the mirror has a log-polar sensor arrangement, such as cameras studied by Tistarelli and Sandini [505] and Pardo et al. [392].

All these works lead to non-central cameras. Nagahara et al. proposed a catadioptric system that is central and has uniform angular resolution [364]; this was possible by using two appropriately shaped and placed curved mirrors, whereas the above works used a single mirror each.

Hicks and Bajcsy as well as Kweon et al. showed how to compute mirror shapes such that a particular scene plane that is "fronto-parallel" to the mirror is imaged without distortion, while still providing a very wide field of view [222, 223, 297]. This was done for both, perspective and orthographic cameras looking at the mirror. Srinivasan contributed a similar result, i.e., a mirror that gives a wide field of view while directly providing a rectified image: in [457], he showed how to compute the shape of a mirror that directly provides cylindrical panoramas, i.e., rectangular images where the two coordinates relate to azimuth respective elevation angles of points relative to the camera's optical axis. He found that there is no smooth mirror shape achieving this property, but that this can be approximated

by a piecewise planar mirror with a layout similar to a Fresnel lens array.

Kondo et al. proposed an anisotropic mirror shape, i.e., that is not a surface of revolution, with the aim of obtaining panoramic vision while allocating higher spatial resolution to a preferred azimuth range, e.g., corresponding to the driving direction of a robot [284, 286].

Nayar and Karmarkar showed how to acquire 360×360 mosaics by stitching together image slices acquired by a rotating *slice camera* [371]. The slice camera is designed such as to have a 360° field of view in one direction, while being orthographic in the other direction. This is achieved by a specially designed mirror; in case the camera looking at the mirror is orthographic itself, the mirror is a cone. Nayar and Karmarkar's design extends slit imaging, cf. Section 2.1.1, to the acquisition of a full spherical field of view, by being based on an omnidirectional slit camera that rotates.

Peleg et al. showed how to acquire circular non-central mosaics using a mirror of a special shape [395, 396], see Section 2.1.1.

A few more general approaches exist for mirror design, as follows. Gaspar et al. proposed a general approach allowing to derive several of the above mirror designs in a unified framework [156, 157]. Hicks as well as Menegatti formulated and provided a solution to the so-called prescribed projection problem [221, 341]: the input is a desired mapping between an object surface and the image plane of a camera in a given position. The goal is to compute a mirror and its location, that together realize this mapping, i.e., the image taken by the resulting catadioptric system is as specified. In many cases, there is no exact solution to the problem, but approximate solutions can be found. Swaminathan et al. addressed the same problem as Hicks in [486] and proposed a solution that minimizes reprojection errors in the image plane, i.e., where the desired and the actual scene-to-image mapping (the "inverse" of the prescribed projection problem) give image points as close to one another as possible. Kondo et al. proposed another approach for the same problem, akin to photometric stereo [285]. This approach allows to conceive discontinuous mirror shapes.

All the above sensors are necessarily non-central (with the exception of [364]), although for many applications, one may model them

sufficiently well using a central approximation, see also the discussion in Section 3.6.

Krishnan and Nayar recently proposed a catadioptric camera whose main optics is a fisheye, a camera they call cata-fisheye [289]. The motivation for this design is to achieve a spherical field of view in azimuth while not necessarily having a hemispherical one in elevation. Indeed, in many applications, such as videoconferencing and vehicle-mounted computer vision, a vertical field of view of a few dozens of degrees is often sufficient. Krishnan and Nayar proposed to achieve this by putting a convex mirror in front of a fisheye. The produced image consists of two annular regions: the outer one corresponds to the direct view at the scene through the fisheye, whereas the inner one shows the reflection in the mirror. Here, the outer image region would usually correspond to the part of the vertical field of view that is above the camera, while the inner region shows a part that is below. One advantage of this system is thus that the full image resolution is dedicated to the desired vertical field of view (as opposed to "wasting" resolution on a full hemispherical fisheye field of view if parts thereof are not relevant for a given application). Other advantages are a good image quality since a mirror of low curvature may be used, and a rather compact design. It is furthermore easy to regulate the desired vertical field of view by displacing the mirror or employing mirrors of different shapes. Strictly speaking, the system is non-central in general, but the working range in which the parallax is negligible, is explained in [289].

2.3.4 Single-Lens Stereo Systems

As mentioned at the beginning of Section 2.3, we consider here catadioptric systems that are intentionally non-central in order to enable stereovision, whereas the systems in the previous section are non-central "by accident" or due to conforming to design goals prohibiting a single effective viewpoint. Most catadioptric single-lens stereo systems are achieved by setting one or more planar or curved mirrors in front of the camera [73, 74, 99, 143, 183, 203, 255, 333, 366, 386, 455] or in addition by keeping a direct view on the scene for a part of the camera's field of view [15, 49, 261, 295, 343, 422, 569].

Advantages of mirror-based stereo systems over multi-camera setups are that no camera synchronization and radiometric alignment of camera are required and that only one set of intrinsic parameters has to be calibrated, as opposed to calibrating each camera in a multi-camera setup.

Mouaddib et al. proposed a set of performance criteria to compare different single-lens catadioptric stereo systems [354, 355].

In the following, we describe some systems, first some using planar mirrors.

Planar mirrors. One of the earliest references suggesting a design for a mirror-based single-lens stereo system is probably a paper of 1899 by Finsterwalder [144], a seminal paper which summarized various results on camera calibration, pose and motion estimation, epipolar geometry, and even projective reconstruction from uncalibrated images. On pages 20–22 of this paper (written in German), Finsterwalder proposed to use a setup of three mutually perpendicular planar mirrors to photograph small objects put between the camera and the mirrors. He suggested that this gives up to eight perspective images (direct view up to multiple reflections) which can be used for stereo reconstruction. We do not know if this system has ever been built.

The idea of using a single planar mirror to acquire stereo images for 3D modeling, a direct and a reflected view of an object, has been considered by various researchers in the early twentieth century, see for example [39, 117, 409, 562, 563]. Some of these works were motivated by the modeling of coastal landscapes from images taken aboard a ship, the mirror surface being formed by a sea or lake. Note also that systems composed of planar mirrors were proposed for stereo *viewing* (as opposed to stereo imaging) as early as in 1838, by Wheatstone [526, 527] (see also [229]).

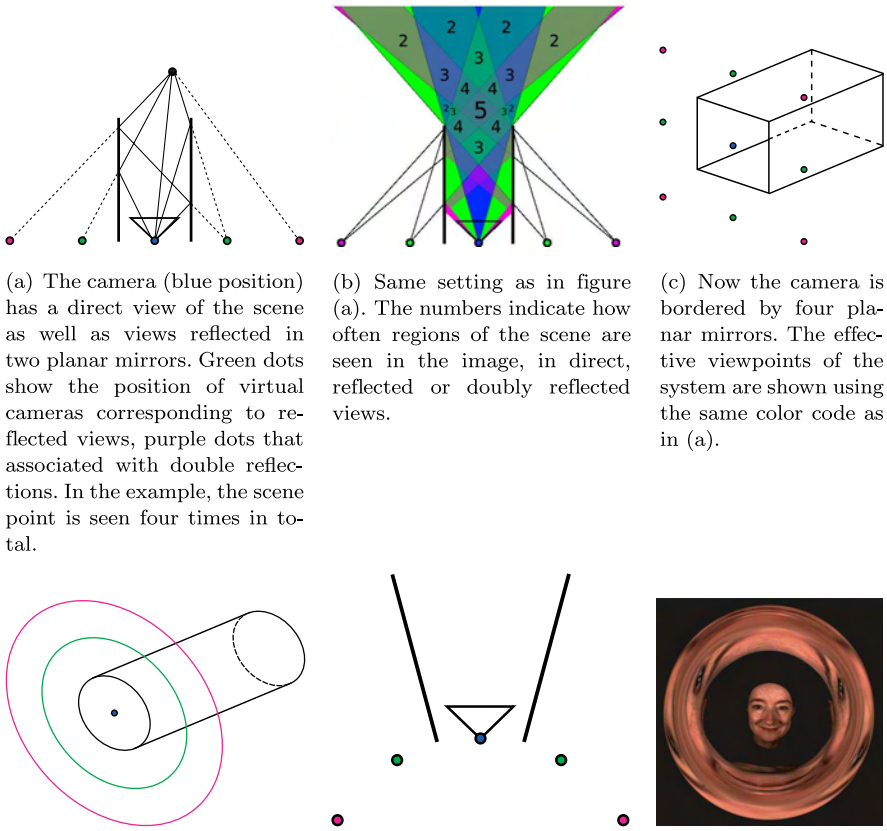
Kaneko and Honda showed that when acquiring an image of an object consisting of a part with a direct view of the object and another with its reflection in a planar mirror, this indeed corresponds to a stereo configuration and allows to reconstruct the object in 3D [261] (see also a similar approach by Zhang and Tsui [569]). This approach was extended by Mitsumoto et al. whose approach relaxed the requirement

of knowing the mirror's position and who also proposed to use an additional planar mirror, to better recover occluded object parts [343]. Arnspang et al. also proposed to put planar mirrors in front of the camera [15, 422]. One of the setups they suggested consisted of two mirrors facing each other and which are parallel to the camera's optical axis (cf. Figure 2.5(a) and (b)). Here, a scene point may be seen multiple times: in the direct camera view, reflected in the mirrors or even after a double reflection, once in each mirror. Arnspang et al. formulated the triangulation problem for their setup and also proposed to extend the setup by arranging more than two mirrors in a cylindrical ring in front of the camera (cf. Figure 2.5(c)).

A similar idea was suggested by Han and Perlin who proposed a single-lens stereo system akin to kaleidoscopes [203]. The camera looks through a tapered tube whose inside is made of planar mirror facets. Each planar mirror, together with the actual camera, corresponds to a virtual camera. Hence, each scene point can be seen multiple times in what is effectively a multi-view stereo system. Furthermore, multiple reflections of light rays may happen inside the tube, thus multiplying the number of virtual cameras and viewpoints, much like what one can observe in a kaleidoscope. Han and Perlin's idea was to use this acquisition system for the acquisition of the Bidirectional Texture Reflectance of objects, based on the large number of viewpoints contained in a single snapshot. Kuthirummal and Nayar proposed a similar system, see further below.

Goshtasby and Gruver used two planar mirrors and no direct view of the scene [183]. Cafforio and Rocca, Inaba et al. as well as Mathieu and Devernay used four planar mirrors, where a pair each gives one virtual camera, via the successive reflections in two mirrors (see Figure 2.6); the system is thus basically equivalent to one with two planar mirrors, but allows for an easier setup by more easily avoiding self-reflections of the camera in the mirrors [74, 244, 333]. Gluckman and Nayar analyzed in detail the relative orientation and epipolar geometry of such systems as well as their self-calibration [174, 175].

In [177], Gluckman and Nayar studied the question of how to obtain an optically rectified stereo pair using a single-lens catadioptric system with planar mirrors. They found that an odd number of mirrors is



(a) The camera (blue position) has a direct view of the scene as well as views reflected in two planar mirrors. Green dots show the position of virtual cameras corresponding to reflected views, purple dots that associated with double reflections. In the example, the scene point is seen four times in total.

(b) Same setting as in figure (a). The numbers indicate how often regions of the scene are seen in the image, in direct, reflected or doubly reflected views.

(c) Now the camera is bordered by four planar mirrors. The effective viewpoints of the system are shown using the same color code as in (a).

(d) Camera looking through a cylinder whose inside is specular. The effective viewpoints are now located on circles. Each viewpoint only "sees" within the plane spanned by it and the axis of the cylinder.

(e) When using a cone instead of a cylinder, the effective viewpoints are no longer coplanar, shown in 2D here.

(f) Image obtained with (e), courtesy of Sujit Kuthirummal [295]. The inner part shows the direct view of the scene, the outer part the reflection in the cone.

Fig. 2.5 Illustration of some single-lens catadioptric stereo systems.

required and for the cases of one and three mirrors, derived the constraints on mirror placement that ensure rectified images. They also showed how to optimize the mirror placement in order to minimize the overall size of the entire sensor.

Avni et al. built a system composed of two cameras and two planar mirrors: the cameras as well as an object to be modeled in 3D are

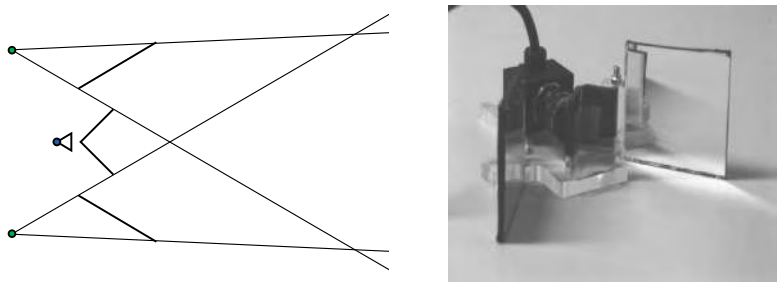


Fig. 2.6 *Left*: Single-lens stereo system using planar mirrors. Here, four mirrors are used for practical reasons, leading to two effective viewpoints, shown in green. *Right:* A practical realization of the system, image courtesy of Frédéric Devernay [333].

positioned above the mirrors such that each camera sees the object's reflections in both mirrors [20]. Hence, a total of four images of the object are obtained, allowing a multi-view 3D modeling.

Non-planar mirrors. Nayar used two specular spheres in the field of view of a camera to obtain stereo information [366]. Southwell et al. presented a design of a single-lens catadioptric stereo sensor with curved mirrors where one convex mirror is fixed on top of a second, larger one [143, 455]. Hence, points in the common field of view of the two individual catadioptric images are seen twice and can be reconstructed. Similar designs were proposed by Cabral et al. [73] and Jang et al. [255]. Jang et al. especially discussed how to maximize the effective stereo baseline.

Nene and Nayar described several single-lens catadioptric stereo configurations where both (virtual) cameras in a stereo system are central [374]. The first uses, like in other systems, two planar mirrors; Nene and Nayar described the epipolar geometry of this setup. The other proposed configurations use mirrors of revolution of conic-shape placed such that the camera is at a focus point of each of the mirrors, see Figure 2.7(a). Hence, each mirror corresponds to a central catadioptric system and by using two or more mirrors arranged this way, a stereo system is obtained where each (virtual) camera is central. Nene and Nayar described the cases of pairs of ellipsoidal, pairs of hyperboloidal and pairs of paraboloidal mirrors and derived the epipolar geometry for

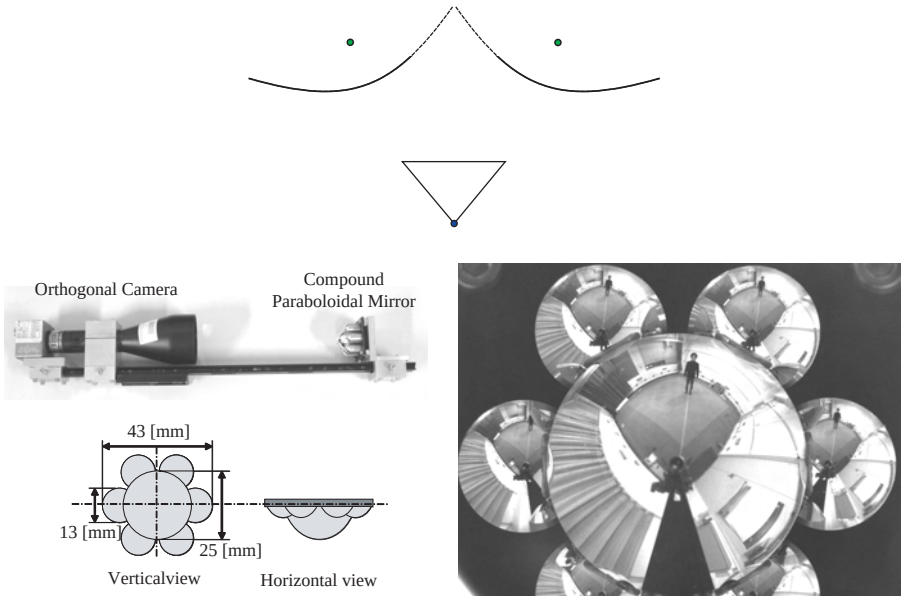


Fig. 2.7 Other single-lens catadioptric stereo systems. *Top*: System using two hyperbolic mirrors with a coinciding focus point, at which the camera is located. This is a single-lens stereo system where each view is central. *Bottom*: Single-lens stereo system by Sagawa et al. composed of paraboloidal mirrors and sample acquired image; the camera looks at the mirrors from below. Images courtesy of Ryusuke Sagawa.

each of those. Other cases are straightforward to imagine, e.g., combing ellipsoidal/hyperboloidal/planar mirrors in the appropriate manner.

Murphy presented a panoramic imaging system for planetary rovers where a single camera looks at a convex mirror with a hole in the middle, through which the camera sees through a fisheye lens [361]. Hence, a single image contains two panoramas, one for the lower part (the catadioptric view) and one for the upper one (fisheye). The system is similar to that of Krishnan and Nayar (cf. Section 2.3.3), although its aim is to obtain single-lens stereo while that of Krishnan and Nayar's design is omnidirectional viewing with a desired distribution of spatial resolution across the image. Benosman et al. proposed a sensor where a camera has both, a direct view of the scene and a view on a hyperbolic mirror [49]. The motivation is to have a high resolution in a dominant direction via the direct perspective image, e.g., the direction ahead of a robot, together with a lower-resolution panoramic view

of the surroundings. Yi and Ahuja proposed another single-lens stereo system, consisting of a hyperboloidal mirror and a concave lens [548]. The camera has a direct view of the mirror and an indirect one of it, through the concave lens, thus effectively producing a stereo pair.

Sagawa et al. built a single-lens catadioptric stereo system consisting of one camera looking at seven spherical or paraboloidal mirrors [432], see Figure 2.7(b) and (c). Although in principle such a system can be used for multi-baseline stereo, the prototype shown in [432] is of small dimension and thus has a small baseline. Its intended application is the detection of close-by objects, for which accurate 3D reconstruction is not necessary. The calibration of the sensor is discussed in [282], see also Section 5.3.2.

Lanman et al. built a similar catadioptric acquisition system consisting of a single high-resolution camera looking at a set of spherical mirrors arranged on a plate [298]. Their system is larger than Sagawa's since one intended application is 3D modeling. The system is obviously non-central since already a single spherical mirror leads to a non-central image. While a single spherical mirror is only "slightly non-central" (more on this in Section 3.6), the setup by Lanman et al. effectively generates a large "baseline", allowing for single image 3D reconstruction. A practical method for calibrating the system, including the properties of second surface mirrors, is proposed in [298], and the multi-mirror view geometry is analyzed, i.e., conditions that hold for points observed in different mirrors, to be images of the same scene point.

Kuthirummal and Nayar proposed a system similar to the one by Arnspang et al. and Han and Perlin (see above), where the piecewise planar tube is replaced by a cone or cylinder-shaped one [295], see Figure 2.5(d)–(f). They demonstrated the use of their system for single-image 3D scene and reflectance recovery.

Orghidan et al. designed a structured-light type depth sensor composed of a camera looking at two hyperboloidal mirrors and a laser emitter [388]. They showed how to model and calibrate this system and use it for 3D modeling.

Rotating mirrors. Other catadioptric single-lens stereo systems have been proposed, using mirrors that rotate between different image

acquisitions in order to produce stereo images, e.g., by Teoh and Zhang [497], Nishimoto and Shirai [376], Murray [362], and Gao and Ahuja [154].

2.3.5 Programmable Systems

Hicks and Nayar et al. proposed methods and system designs allowing to modify the shape of the mirror in a catadioptric system in order to adapt the sensor to a change in the scene or to acquire an image in a desired way [224, 296, 370].

Nayar et al. proposed to use a programmable array of micro-mirrors to achieve programmable, or purposive, imaging [370, 369]. They demonstrated the use of digital micro-mirror devices (DMDs), routinely used in projectors, for tasks such as high dynamic range image acquisition, optical appearance matching, or generally speaking, the change of imaging geometry. In a similar work, Hicks et al. also suggested to use micro-mirrors to control the way an image is acquired, e.g., by actively controlling the spatial resolution across an image [224].

The following two ideas are not strictly speaking programmable systems. Kuthirummal and Nayar studied the possibility of using a flexible mirror sheet as reflector in a catadioptric system [296]. They proposed to recover the current shape of the mirror from its outline in the image and additional assumptions; once the shape is determined, the catadioptric system is effectively calibrated. The system is not programmable in the same sense as those by Hicks and Nayar et al. but is described in this section since it operates by changing the mirror shape during image acquisition.

Fergus et al. proposed the concept of random lens imaging [140]. One practical instance of this concept is a camera looking at a collection of randomly positioned small mirrors or refractive elements. A calibration method is proposed and potential applications of the general concept for tasks, such as super-resolution and depth sensing, are described.

2.4 Stereo and Multi-camera Systems

Using two or more cameras to achieve omnidirectional viewing is a well-known principle and will not be covered in great detail here, besides a

few historical references and references to stereo systems based on omnidirectional cameras. A nice overview of omnidirectional stereo systems is due to Zhu [573].

Lin and Bajcsy proposed a catadioptric stereo system consisting of two cameras looking at one cone-shaped mirror from different distances [321]. The cameras are aligned with the cone's axis of revolution; to be precise, one camera lies on the axis, the other one looks at the mirror through a beam splitter such that it virtually comes to lie on the mirror axis. Hence, epipolar planes contain the axis, which simplifies the epipolar geometry and 3D reconstruction.

Spacek proposed to use two cone–camera pairs, one on top of the other and all cameras and mirrors axis-aligned [456]. He derives the equations for stereo computations and highlights that cone-based catadioptric cameras, although being non-central, can be used for stereo and may be advantageous over the more common central catadioptric systems due to their higher spatial resolution (lower vertical field of view). A similar system, consisting of two hyper-catadioptric cameras, was used in [57, 413].

Other stereo systems combine omnidirectional and traditional cameras, to combine their respective advantages, see e.g., [3, 79, 118, 268].

Multi-camera systems composed of several *perspective* cameras have been built at least as early as in 1884, initially mainly if not exclusively for aerial imaging. The earliest work known to us (no effort was made for an exhaustive bibliography research) is that of Triboulet, who, as reported in [504] experimented from 1884 on with a multi-camera system consisting of seven cameras attached to a balloon: one camera looked downward and six cameras were equally distributed around the balloon's circumference (the system thus resembles the popular Ladybug sensor by Point Grey, `http://www.ptgrey.com`). Similar other multi-camera systems from the early twentieth century include those by Scheimpflug [438] and Thiele [499], cf. Figure 2.8. Several other systems were developed throughout the first half of the twentieth century, see for example [87, 403, 423, 502].

Other systems used multiple lenses in the same camera body. Gasser as well as Aschenbrenner invented systems where multiple cameras share the same focal plane and film; due to a suitable arrangement

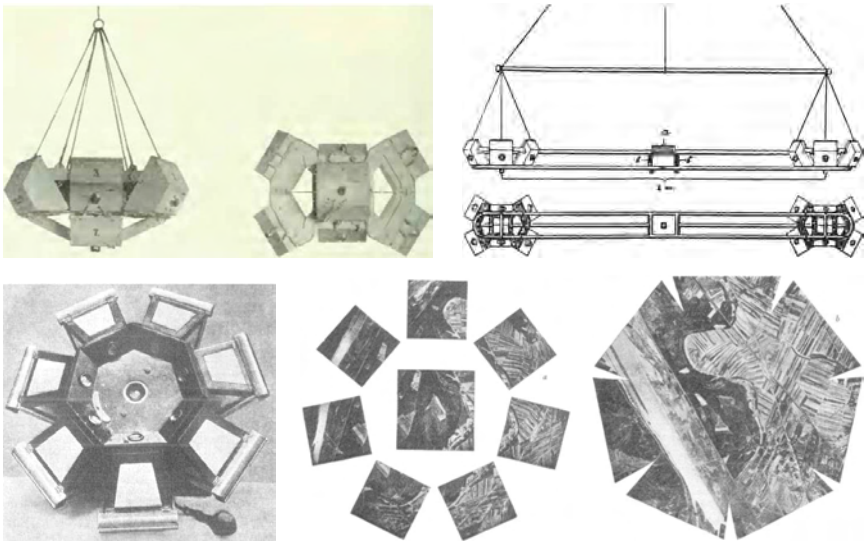


Fig. 2.8 Three multi-camera systems from the first decade of the twentieth century. *Top*: two systems developed by Thiele [499]. The first one, called auto-panoramograph, was attached at the bottom of a balloon. The second system, the stereo-panoramograph, is equivalent to an omnidirectional stereo rig with a baseline of 2 m. *Bottom*: A system developed by Scheimpflug [438], consisting of seven oblique cameras and a central one. Shown are eight original images acquired with the system and a mosaic composed from them using a special projection equipment called photo-perspectograph.

of prisms and mirrors, the different cameras directly generate a composite wide-angle image on the film [18, 160]. These systems were, like most of the above, intended for aerial imagery. Multi-lens cameras were used at least as early as in 1882: Londe used multi-lens cameras to study epileptic seizures, by capturing a succession of time-delayed images [71], similar to the famous motion series by Muybridge according to an acquisition principle he developed in 1872 [71].

2.5 Others

2.5.1 Plenoptic Camera

Adelson and Wang proposed the so-called plenoptic camera, a design that essentially gives multiple pinhole images while using a single main aperture and a single sensor array [2]. When placing a lenticular array in front of a camera's sensor plane, each lenslet together with a subset

of the sensor array's pixels forms a tiny pinhole camera. That pinhole camera effectively only captures rays entering the camera's main aperture within a subregion of the aperture, as opposed to regular cameras, where each pixel integrates light rays from all over the aperture. Conceptually, a plenoptic camera thus corresponds to a multi-camera system consisting of cameras arranged on a planar grid. Acquired images can be used to perform stereo computations or various computational photography tasks. We will not consider plenoptic cameras further in this article since they are not usually used for panoramic imaging, although in principle nothing would prevent to use them with fisheye lenses (besides the drop of spatial resolution for the already low-resolution individual pinhole sub-cameras).

More information on plenoptic cameras and similar designs can for example be found on `http://en.wikipedia.org/wiki/Integral_photography` and in [429, 162] The concept of plenoptic camera is highly related to that of integral imaging [323] and parallax stereogram [248] or panoramagram [249].

2.5.2 Biprism

A single-lens stereo system using a biprism rather than a mirror, that directly produces rectified images, has been developed by Lee et al. [303].

2.5.3 Spherical Lens and Spherical Image Area

Krishnan and Nayar proposed an omnidirectional sensor consisting of a transparent sphere, around which sensing elements are uniformly arranged on a spherical surface, with free space between neighboring elements [290]. Hence, individual sensing elements can see through the transparent sphere and other sensing elements, to perceive the scene. The design goals of the system are an omnidirectional field of view with a single center of projection and uniform spatial resolution.

2.5.4 Krill-Eye

Hiura et al. proposed an optical arrangement for wide-angle imaging system, termed krill-eye, that is inspired by compound animal eyes [228]

The system consists of gradient refractive index (GRIN) lenses aligned on a spherical surface. Hiura et al. proposed a theoretical study on image quality and focusing properties and built a proof-of-concept prototype.

2.5.5 Rolling Shutter Cameras

Meingast et al. and Ait-Aider et al. studied the case of rolling shutter CMOS cameras, where the image area is not exposed simultaneously, but in a rolling fashion across rows of pixels [10, 171, 340]. The consequence is that when taking a picture of a moving object, it will appear distorted. Meingast et al. developed projection models for the general and several special cases of object motion and showed how optical flow can be modeled [340]. They also addressed the calibration of the sensor. Ait-Aider et al. used the rolling shutter's apparent drawback to develop an approach that computes both, the pose of a known object and its instantaneous velocity, from a single image thereof [10]. There is a conceptual link to pushbroom cameras, as follows, highlighted in a similar fashion in [171, 340]. An alternative way of looking at the rolling shutter phenomenon is to consider a static object and a pushbroom camera moving around it with a velocity that combines the inverse velocity of the object and a motion that compensates the difference in pose associated with different rows of pixels in the camera.

The effect of rolling shutters was also taken into account by Wilburn et al. who built a multi-camera array with CMOS cameras [529, 530], with several applications such as synthetic aperture imaging or allowing to acquire videos with increased frame rate, dynamic range, resolution, etc. A method for calibrating such a camera array was proposed by Vaish et al. [516].

3

Camera Models

In the following, we review a certain number of camera models that were proposed in the literature. Some of them are based on physical models, others are of a more algebraic nature.

These models can be described and sorted according to various criteria. A first characteristic concerns the spatial distribution of the camera rays along which a camera samples light in the environment. Most models have a single optical center through which all camera rays pass. We also speak of **central camera models**. For these, the back-projection function (see below) delivers the direction of the camera ray. **Non-central camera models** do not possess a single optical center. In that case, the back-projection operation has to deliver not only the direction but also the position of a camera ray, e.g., some finite point on the ray. We will also use Plücker coordinates to represent camera rays. Special cases of non-central cameras are **oblique camera**, where no two camera rays meet [390] and **axial cameras** where there exists a line that cuts all camera rays. A special case of this are **x-slit** or **two-slit models** where there exist two lines that cut all camera rays. This is for example the case for linear pushbroom cameras [197].

A second property of camera models and one we would like to stress particularly, concerns how global or local a model is. The following

definitions have already been introduced in Section 1.1, cf. also Figure 1.1. Most models have a small set of parameters and in addition, each parameter influences the projection function all across the field of view. We call these, **global models**, since they hold for the entire field of view/image area. Second, there exist several more **local models**, e.g., models with different parameter sets for different portions of the image. These models have usually more parameters than global ones, but have a higher descriptive power. Finally, at the extreme, we have **discrete models**, where the projection or back-projection function is merely sampled at different points, e.g., sampling the back-projection function at every pixel. These models have many parameters, one set of them per sampled location. They are sometimes called non-parametric models, but we do not find this entirely appropriate, since they do have parameters; hence the proposed name of discrete models. Let us note that the transition from global to local to discrete models is not discontinuous: some local models have the same parametric form as global ones, and to use discrete models, one usually needs some interpolation scheme, for example to be able to back-project any image point, not only sampled ones. In that respect, discrete models plus, such as interpolation scheme, are actually local models in our language.

Many of the global and discrete models described in the following are well known. This is less true for the local models, although they may represent a good compromise between tractability (number of parameters, stability of calibration) and generality. We would like to point out the work of Martins et al. [331], which is rather often cited but often described only partially or not appropriately. They proposed three versions of the so-called two-plane model, see Sections 3.1.3 and 3.2.1 for more details. These foreshadowed several important contributions by others, mostly achieved independently. First, it contains one of the first proposals for a local camera model (the spline-based version of the developed two-plane model). Second, it sparked works on discrete camera models, where camera rays are calibrated individually from images of two or more planar calibration grids [85, 189], an approach rediscovered by [191, 477]. Third, the linear and quadratic versions of the two-plane model, when written in terms of back-projection matrices, are nothing else than particular

instances of rational polynomial models; they are thus closely related to the division and rational models explained by Fitzgibbon [145] and Claus and Fitzgibbon [98] or the general linear cameras of Yu and McMillan [559], as well as others, which found a lot of interest recently. Martins et al. may also be among the first to explicitly use lifted coordinates (cf. Section 1.2) in camera models.

Finally, a third main property of camera models we consider is that some models have a direct expression for forward projection, others for back-projections, some work easily both ways. This is important since forward and back-projections are convenient for different tasks: forward projection for distortion correction of images and bundle adjustment, back-projection for minimal methods for various structure-from-motion tasks, such as pose and motion estimation.

A few more general notes and explanations of notations follow.

Back-projection versus forward projection. One defines the other but one or the other may be more difficult to formulate algebraically. For example, the classical radial distortion model is written using polynomials; even with only one distortion coefficient, the polynomials are cubic, making it cumbersome to write the inverse model. This model is traditionally used for back-projection although some researchers used the same expression for forward projection. Other polynomial models are also used for either direction.

Generally speaking though, back-projection is often easier to formulate. This is especially true for catadioptric systems, where back-projection comes down to a deterministic "closed-form" ray-tracing whereas forward projection entails a search for which light ray(s) emitted by a 3D point gets reflected by the mirror(s) into the camera. Also, when using rational polynomial functions in point coordinates to express camera models, this is straightforward for back-projection; using such a model for forward projection, results in general in curved "camera rays", i.e., the set of 3D points mapped onto an image point, form a curve, not a straight line (cf. Section 3.1.8 on the cubic camera model by Hartley and Saxena [210]).

For these and other reasons, we will emphasize back-projection formulations in the following. Another reason for this is that back-projection is less often used in the literature although it enables to

clearly formulate several structure-from-motion tasks and epipolar geometry.

Back-projection will be formulated in two different ways, either by giving expressions allowing to compute two points on a camera ray, or by directly giving the Plücker coordinates of the ray. As for the first case, one of the two points on the ray will be its point at infinity and the other, a finite point, denoted by

$$\mathbf{B}^i \quad \text{and} \quad \mathbf{B}^f,$$

respectively. The point at infinity will be given as a 3-vector, the finite point as a 4-vector, both in homogeneous coordinates. Note that in this survey we do not consider the possibility of camera rays that lie completely on the plane at infinity; such a case is of small practical interest and if necessary, all formulas are easy to adapt. Plücker coordinates of camera rays will be written as

$$\mathbf{B}^l.$$

They can of course be easily computed from the two points $\mathbf{B}^i$ and $\mathbf{B}^f$, cf. Section 1.2.

The reason for using both expressions for back-projection is that the two-point expression is better suited to formulate pose estimation, whereas the Plücker-based expression allows to formulate motion estimation and epipolar geometry in a straightforward manner. When we deal with central camera models, we will only give the point at infinity $\mathbf{B}^i$ of camera rays, i.e., their direction, always assuming that cameras are in canonical position, i.e., with the optical center at the origin.

Whenever possible, we try to write the back-projection operation using back-projection matrices, operating on (lifted) image point coordinates. We will use full back-projection matrices

$$\mathsf{B}^l$$

of size $6 \times n$, that map image points to camera rays in Plücker coordinates, and partial back-projection matrices

$$\mathsf{B}^i$$

of size $3 \times n$, that map image points to camera ray directions. The values of n for different models depend on which liftings of image coordinates are required.

Finally, let us note that the back-projection for a true camera usually gives rise to a half-line in 3D; many camera models do not take this fully into account and only model back-projection via infinite lines. We do the same in this monograph and, for ease of language, use the term ray or camera ray to denote a line of sight of a camera.

Radial distortion. In this section, models are presented in different forms, depending if they are backward or forward ones, or only model radial distortion or more general distortions. Radial distortion models are given in different forms; they are defined by a 1D (un)distortion profile or (un)distortion function, mapping radial distances in the distorted image (distances of image point to distortion center) to either radial distances in the undistorted image or the incidence angle between camera rays and the optical axis.

We use the following notations. The radial distance is denoted by r_d (in distorted images) and r_u (in undistorted images). The angle between a camera ray and the optical axis is denoted by θ. Radial distortion models are usually given as a mapping between any of these three entities. For simplicity, we always assume in the following that the radial distortion center is equal to the principal point and that it is known and located at the origin. When considering a distortion center different from the origin, the given back-projection equations can be modified in a straightforward manner, by preceding them with the translation bringing the distortion center to the origin. When considering a distortion center different from the principal point, the modifications of back-projection equations are equally straightforward. Whereas in the following, we suppose that the perspective back-projection consists merely in "undoing" the focal length, considering a distortion center that is different from the principal point, requires a full perspective back-projection.

Extrinsic parameters. Camera pose, or extrinsic parameters, is straightforward to take into account. As for back-projection, translations and rotations can be applied as follows (cf. Section 1.2):

$$\mathbf{B}^f \mapsto \begin{pmatrix} \mathsf{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{B}^f$$

$$\mathbf{B}^i \mapsto \mathsf{R}\mathbf{B}^i$$
$$\mathbf{B}^l \mapsto \begin{pmatrix} \mathsf{R} & 0 \\ -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \end{pmatrix} \mathbf{B}^l.$$

3.1 Global Camera Models

3.1.1 Classical Models

By "classical models", we mean those used most often in applications and academic research, i.e., the pinhole model, affine camera models, and pinhole models enhanced with classical terms for radial and tangential distortions.

Pinhole model. The pinhole model, or perspective projection, assumes that all camera rays pass through a single point, the optical center and that there is a linear relationship between image point position and the direction of the associated camera ray. That relationship can be expressed via a so-called calibration matrix which depends on up to five intrinsic parameters:

$$\mathsf{K} = \begin{pmatrix} f_u & s & x_0 \\ 0 & f_v & y_0 \\ 0 & 0 & 1 \end{pmatrix},$$

where f_u respectively f_v is the focal length measured in pixel dimensions, horizontally respectively vertically, s is a so-called skew term which may for example model non-rectangular pixels or synchronization errors in the image read-out, and (x_0, y_0) are the coordinates of the principal point, the orthogonal projection of the optical center onto the image plane. The linear relation between image points and ray directions can be written as:

$$\mathbf{q} \sim \mathsf{K}\mathbf{B}^i.$$

The model is equally easy to use for forward and back-projections. In what follows, we will assume square pixels, i.e., $s = 0$ and $f_u = f_v = f$:

$$\mathsf{K} = \begin{pmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3.1}$$

Affine models. We also briefly mention affine camera models, where the optical center is a point at infinity. Various submodels thereof exist, often called orthographic, weak-perspective, and para-perspective models, going from the most specific to the most general affine model [13, 232]. These models can give good approximations of pinhole cameras in the case of very large focal lengths or if the scene is shallow in the direction of the camera's optical axis as well as far from the camera. Back-projection is slightly different from the pinhole model in that the direction is the same for all camera rays (the direction corresponding to the optical center) and the essential back-projection operation is thus not the computation of the ray direction, but of a finite point on the ray.

Classical polynomial distortion models. The pinhole model has been enhanced by adding various terms for radial, tangential, and other distortions. The most widely established models are described for example in [63, 64, 149, 152, 451]. The usual model for combining radial and decentering distortion (one type of tangential distortion) is:

$$\begin{aligned}
\bar{x}_d &= x_d - x_0 \\
\bar{y}_d &= y_d - y_0 \\
r_d^2 &= \bar{x}_d^2 + \bar{y}_d^2 \\
x_u &= x_d + p_1(r_d^2 + 2\bar{x}_d^2) + 2p_2\bar{x}_d\bar{y}_d + \bar{x}_d \sum_{i=1}^{n} k_i r_d^{2i} \\
y_u &= y_d + 2p_1\bar{x}_d\bar{y}_d + p_2(r_d^2 + 2\bar{y}_d^2) + \bar{y}_d \sum_{i=1}^{n} k_i r_d^{2i},
\end{aligned} \tag{3.2}$$

where n is the number of radial distortion coefficients used. The distortion corrected image points correspond to a pinhole model, i.e., are related to camera ray directions by:

$$\mathbf{B}^i \sim \mathsf{K}^{-1} \begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix}.$$

In the above model, it is assumed that the radial distortion center coincides with the principal point. Other models exist that consider them as distinct and/or include additional distortion terms, see e.g., [149, 215].

If we neglect tangential distortion and suppose that image coordinates are centered in the distortion center and that the latter is equal to the principal point, then we can deduce, from Equation (3.2), the radial (un)distortion function:

$$r_u = r_d \left(1 + \sum_{i=1}^{n} k_i r_d^{2i} \right). \tag{3.3}$$

This relates the distance of a point from the distortion center, before and after distortion. From this, it is easy to extract the relationship between the radial distance in the distorted image and the incidence angle of the camera ray with the optical axis:

$$\theta = \operatorname{atan} \frac{r_u}{f} = \operatorname{atan} \frac{r_d \left(1 + \sum_{i=1}^{n} k_i r_d^{2i} \right)}{f}, \tag{3.4}$$

which depends on the focal length of the camera.

Let us now consider a radial distortion with a single coefficient k_1. We can write down the camera ray direction in the following form:

$$\begin{aligned} \mathbf{B}^i &\sim \mathsf{K}^{-1} \begin{pmatrix} x_d + k_1 \bar{x}_d r_d^2 \\ y_d + k_1 \bar{y}_d r_d^2 \\ 1 \end{pmatrix} \sim \begin{pmatrix} (x_d - x_0)(1 + k_1 r_d^2) \\ (y_d - y_0)(1 + k_1 r_d^2) \\ f \end{pmatrix} \\ &\sim \mathsf{B}^i_{3\times 10} \mathcal{L}^3(\mathbf{q_d}), \end{aligned}$$

where the lifting operation $\mathcal{L}^3(\cdot)$ is given in Equation (1.4) and the back-projection matrix is:

$$\mathsf{B}^i \sim \begin{pmatrix} k_1 & 0 & k_1 & 0 & -3k_1x_0 & -2k_1y_0 & -k_1x_0 & 1+k_1a & 2k_1x_0y_0 & -x_0(1+k_1c) \\ 0 & k_1 & 0 & k_1 & -k_1y_0 & -2k_1x_0 & -3k_1y_0 & 2k_1x_0y_0 & 1+k_1b & -y_0(1+k_1c) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & f \end{pmatrix},$$

with $a = 3x_0^2 + y_0^2, b = x_0^2 + 3y_0^2$ and $c = x_0^2 + y_0^2$.

Since the first and third columns are equal, as well as the second and fourth ones, we may use a more compact back-projection matrix of size 3×8:

$$\mathsf{B}^i_c \sim \begin{pmatrix} k_1 & 0 & -3k_1x_0 & -2k_1y_0 & -k_1x_0 & 1+k_1a & 2k_1x_0y_0 & -x_0(1+k_1c) \\ 0 & k_1 & -k_1y_0 & -2k_1x_0 & -3k_1y_0 & 2k_1x_0y_0 & 1+k_1b & -y_0(1+k_1c) \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & f \end{pmatrix},$$

which operates on lifted image coordinate vectors of the form:

$$\begin{pmatrix} q_1^3 + q_1 q_2^2 \\ q_1^2 q_2 + q_2^3 \\ q_1^2 q_3 \\ q_1 q_2 q_3 \\ q_2^2 q_3 \\ q_1 q_3^2 \\ q_2 q_3^2 \\ q_3^3 \end{pmatrix}.$$

If we assume a known principal point (equal to the distortion center here) and translate image points such that the principal point lies at the origin, then the back-projection equations reduce further to:

$$\mathsf{B}_c^i \sim \begin{pmatrix} k_1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & k_1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & f \end{pmatrix},$$

which can be compacted similarly to the above:

$$\mathsf{B}_{cc}^i \sim \begin{pmatrix} k_1 & 0 & 1 & 0 & 0 \\ 0 & k_1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & f \end{pmatrix} \quad \text{operating on} \quad \begin{pmatrix} q_1^3 + q_1 q_2^2 \\ q_1^2 q_2 + q_2^3 \\ q_1 q_3^2 \\ q_2 q_3^2 \\ q_3^3 \end{pmatrix}.$$

Back-projection matrices for combined radial and tangential distortion models are equally straightforward to derive. Note that this back-projection formulation allows to easily write down equations for the epipolar geometry, cf. Section 4.2.1.

Inversion of models. Devernay and Faugeras discussed the analytical inversion of the classical radial distortion model with one coefficient, using Cardan's method. This analysis was adapted by Ma et al. to other variants of polynomial radial distortion models, especially rational polynomials [327], cf. also Section 3.1.8. Heikkilä showed how to formulate an approximate inverse model for radial and tangential distortions [215]. A more detailed and accurate inverse model was proposed by Mallon and Whelan [330].

Other models and works. The standard radial and tangential distortion model is a backward one, going from distorted to undistorted coordinates. Often, the same model is also employed for forward projection, see e.g., [41, 216, 525, 521], in which case the model coefficients' values will differ of course.

Many researchers have used other polynomial models for distortion/undistortion, e.g., the so-called polynomial fisheye transform (PFET) by Basu and Licardie [41], which is nothing else than a general polynomial relating r_d and r_u, including even terms. Such models were for example applied for endoscope calibration, see [17, 204, 454, 565]. Other calibration approaches using such models are [8, 42, 134, 148, 215, 216, 274, 302, 330, 381, 436, 437, 514, 515, 521, 525, 534]. These approaches often differ with respect to the choice of calibration object, potentially carrying out controlled motions to position it in front of the camera, and with respect to the strategy for estimating the unknowns (e.g., estimating some parameters first while keeping others at initial values).

Kölbl proposed to use a trigonometric series to model radial distortion, in order to better condition the normal equations occurring in distortion estimation [283]. Several other distortion models are described in [275, 276].

Shih et al. established an approach for evaluating the errors made when neglecting lens distortion during calibration [449]. This is a function of the actual amount of true distortion, the number and positions/distribution of calibration points, the amount of noise and possibly other factors.

Very wide fields of view. It is well known that the above classical radial distortion model is not appropriate for cameras with a very wide field of view. To see this, refer to Equation (3.4): in order to achieve a hemispheric field of view ($\theta = 90°$), one would need r_d and thus the extent of the image area, to be infinitely large. Many alternative models exist, described in subsequent sections.

3.1.2 Extensions for Modeling Zooming and Focusing

In [64], Brown explained that lens distortion varies with a camera's focus setting but that knowledge of the distortion for two focal settings

allows to interpolate the distortion for any other setting. Also, radial distortion depends on the distance of the object to the camera. These effects are rigorously proven but seem to be used relatively seldomly in practice.

There exist various approaches for modeling zoom lenses and the influence of zooming on the intrinsic parameters, as well as for modeling other intricate phenomena related to camera calibration, see e.g., [11, 16, 53, 70, 96, 100, 102, 128, 150, 313, 469, 531, 533, 532].

3.1.3 Two-Plane Model

Chen et al. introduced implicit camera models in computer vision [89, 90]. Their two-plane model is mostly known through the subsequent paper by Martins et al. [331] and follow-up papers by Wei and Ma [522, 524] and others. As explained in the introduction of this section, the two-plane model is rather rich in ideas, which is why it will be covered in some detail here.

The model is defined as follows [331]. Consider two images of a planar calibration grid, put in known positions and assume that we can extract matches between the image and grid in both images. Since the grid positions are known, points on the grid actually define a 3D calibration object, which may be used as input to classical calibration approaches. The motivations for the two-plane model are to not rely on pre-defined camera models for the calibration and to handle non-central cameras. This may be achieved as follows. If we have dense matches, we may compute, for each image point for which we have two matches, the associated camera ray, simply by computing the line spanned by the two matched grid points. This idea was fully introduced later, by Gremban et al. [189] and Grossberg and Nayar [191]. Martins et al. did not assume dense matches; hence, in order to compute camera rays, one must have recourse to some interpolation scheme. Basically, for an image point we want to back-project, we need to determine the corresponding points on the two grids. To do so, Martins et al. proposed three interpolation schemes to compute corresponding points on calibration grids, from extracted matches: two global schemes (linear

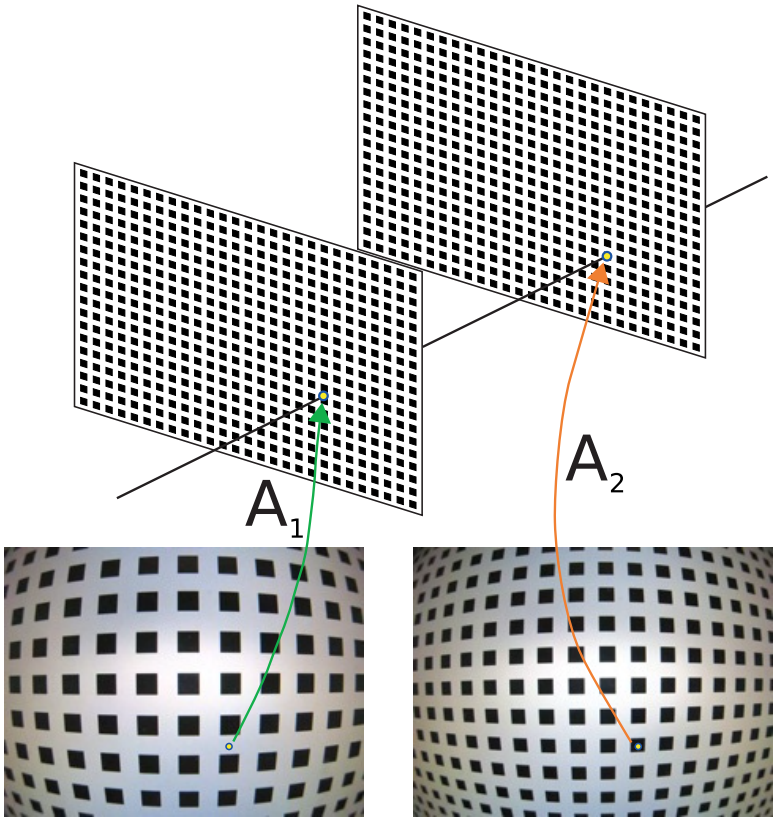


Fig. 3.1 Illustration of the two-plane model. A_1 and A_2 represent the interpolation transformations for the two grids, affine respectively quadratic transformations for the linear respectively quadratic versions of the model.

and quadratic) and a local one (spline-based). The global schemes are illustrated in Figure 3.1.

First version — "Linear interpolation". The mapping between the image plane and each planar calibration grid was supposed to be an affine transformation:

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \mathsf{A}_i \, \mathbf{q},$$

where $i = 1, 2$ is the index of the grid and

$$\mathsf{A}_i = \begin{pmatrix} A_{i,11} & A_{i,12} & A_{i,13} \\ A_{i,11} & A_{i,12} & A_{i,13} \\ 0 & 0 & Z_i \end{pmatrix},$$

and it is assumed that $q_3 = 1$. Here, Martins et al. assumed that the two calibration grids are parallel to the X, Y plane, at known depths Z_i. They stated that other cases can also be handled similarly, but did not develop this further.

A more proper formulation using homogeneous coordinates also for 3D points, would separate the pose of the grid, parameterized by a rotation and translation, from the interpolation function "inside" the plane, which can be an affine transformation (as in Martins et al.'s work) or a projective one:

$$\mathbf{Q}_i \sim \underbrace{\begin{pmatrix} \mathsf{R}_i & \mathbf{t}_i \\ \mathbf{0}_3^\mathsf{T} & 1 \end{pmatrix}}_{\text{pose matrix}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\mathsf{M}} \mathsf{A}_i \, \mathbf{q} \sim \begin{pmatrix} \bar{\mathsf{R}}_i & \mathbf{t}_i \\ \mathbf{0}_2^\mathsf{T} & 1 \end{pmatrix}_{4\times 3} \mathsf{A}_i \, \mathbf{q},$$

where $\bar{\mathsf{R}}_i$ consists of the first two columns of R_i and M maps 2D coordinates on the grid plane, to 3D coordinates. The computation of the interpolation mappings A_i (i.e., the **calibration**) would be carried out between image points and 2D points on calibration grids, instead of 3D points as in the original formulation above.

The back-projection ray of an image point $\mathbf{q}$ can be computed from the two grid points $\mathbf{Q}_1$ and $\mathbf{Q}_2$ according to Equation (1.1) in Section 1.2. Let $\mathsf{P}_i = \begin{pmatrix} \bar{\mathsf{R}}_i & \mathbf{t}_i \end{pmatrix}$. Then:

$$\mathbf{B}^l \sim \begin{pmatrix} (\mathsf{A}_2\mathbf{q})_3 \mathsf{P}_1 \mathsf{A}_1 \mathbf{q} - (\mathsf{A}_1\mathbf{q})_3 \mathsf{P}_2 \mathsf{A}_2 \mathbf{q} \\ (\mathsf{P}_1 \mathsf{A}_1 \mathbf{q}) \times (\mathsf{P}_2 \mathsf{A}_2 \mathbf{q}) \end{pmatrix}.$$

This expression is quadratic in the coordinates of the image point $\mathbf{q}$. It is thus clear that back-projection can be written via a 6×6 back-projection matrix:

$$\mathbf{B}^l \sim \mathsf{B}^l_{6\times 6} \, \mathcal{L}^2(\mathbf{q}) \tag{3.5}$$

operating on lifted image point coordinates $\mathcal{L}^2(\mathbf{q})$ (cf. Equation (1.4)).

In the special case of calibration planes parallel to the XY-plane and of affine transformations A_i, we get:

$$\mathsf{B}^l = \begin{pmatrix} 0 & 0 & 0 & A_{1,11} - A_{2,11} & A_{1,12} - A_{2,12} & A_{1,13} - A_{2,13} \\ 0 & 0 & 0 & A_{1,21} - A_{2,21} & A_{1,22} - A_{2,22} & A_{1,23} - A_{2,23} \\ 0 & 0 & 0 & 0 & 0 & Z_1 - Z_2 \\ 0 & 0 & 0 & Z_2 A_{1,21} - Z_1 A_{2,21} & Z_2 A_{1,22} - Z_1 A_{2,22} & Z_2 A_{1,23} - Z_1 A_{2,23} \\ 0 & 0 & 0 & Z_1 A_{2,11} - Z_2 A_{1,11} & Z_1 A_{2,12} - Z_2 A_{1,12} & Z_1 A_{2,13} - Z_2 A_{1,13} \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \end{pmatrix}, \tag{3.6}$$

where the last row is given by the vector

$$\mathbf{b} = \begin{pmatrix} A_{1,11}A_{2,21} - A_{1,21}A_{2,11} \\ A_{1,11}A_{2,22} + A_{1,12}A_{2,21} - A_{1,21}A_{2,12} - A_{1,22}A_{2,11} \\ A_{1,12}A_{2,22} - A_{1,22}A_{2,12} \\ A_{1,11}A_{2,23} + A_{1,13}A_{2,21} - A_{1,21}A_{2,13} - A_{1,23}A_{2,11} \\ A_{1,12}A_{2,23} + A_{1,13}A_{2,22} - A_{1,22}A_{2,13} - A_{1,23}A_{2,12} \\ A_{1,13}A_{2,23} - A_{1,23}A_{2,13} \end{pmatrix}.$$

If instead of affine transformations, projective ones are adopted for A_1 and A_2, then the back-projection matrix B^l has no zero coefficients anymore.

Second version — "Quadratic interpolation". Here, the interpolation function between the image plane and each of the calibration grids, is assumed to be quadratic, and modeled by:

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \mathsf{A}_i \, \mathcal{L}^2(\mathbf{q}),$$

with 3×6 matrices A_i. Obviously, camera rays can be expressed via a back-projection matrix operating on degree-4 liftings of the image coordinates:

$$\mathbf{B}^l \sim \mathsf{B}^l_{6\times 15} \, \mathcal{L}^4(\mathbf{q}).$$

Third version — "Linear spline interpolation". This is a local camera model according to our definition and will thus be explained in Section 3.2.1.

Observations. In [331] it is noted that "The basic difference between this model and the pinhole model is that all lines of sight are not forced to go through the same point". A second major difference is that this model allows to handle non-perspective distortions that may in principle be quite general.

The linear two-plane model is actually identical to the general linear camera (GLC) model of Yu and McMillan [559], cf. Section 3.4.4, although it is formulated differently. Also, whereas the two-plane model was proposed in order to interpolate calibration data, the GLC was introduced in order to generate novel spatial distributions of camera rays, with applications for example in image synthesis.

The linear and quadratic two-plane models are particular instances of rational polynomial camera models, which were later introduced in full generality, cf. Section 3.1.8. Also, the two-plane models may be among the first to explicitly use lifted image coordinates in their formulation.

Finally, as mentioned above, they inspired works on discrete camera models; with the possibility of getting dense matches, it is indeed natural to calibrate camera rays directly for sampled image positions instead of computing interpolation transformations, as proposed by Gremban et al. [189], Grossberg and Nayar [191], and others, see Section 3.3.

Extensions. Izaguirre et al. showed how to incorporate the camera motion in the back-projection matrix [251]. Since we have explicitly written the back-projection, this can also be done easily by combining back-projection (e.g., Equation (3.5) for the linear two-plane model) with extrinsic parameters, as explained in Equation (1.2).

The two-plane model is formulated as a back-projection; Gremban et al. proposed an approach to deduce the forward projection operation [189]. In addition, they already posed the basis for a fully ray-based camera calibration, which is treated in Section 3.3.

The two-plane model was further generalized by Wei and Ma in several articles, as follows. In [522] necessary and sufficient constraints were derived for the back-projection matrix to correspond to a perspective back-projection, i.e., when no non-perspective distortion is present. In [523, 524] the authors also modeled the forward projection, using

the same algebraic model but with independent parameters and estimation processes. There, they also investigated the epipolar geometry associated with their camera model. The findings of [524] do not seem to be fully general since both for the forward projection and the epipolar geometry, the existence of an optical center of the cameras seems to be assumed, whereas the considered camera model in general corresponds to non-central cameras.

It would be worthwhile to establish the conditions under which two cameras following the two-plane model, form a stereo pair, along the lines of Seitz and Kim's study [445], see Section 3.4.2. Further, it would be interesting to establish necessary and sufficient constraints on the back-projection matrices under which the camera is central.

Champleboux et al. extended the two-plane model in two respects [85]. First, by allowing the use of more than two images of calibration planes. Second, by modeling the function that maps image points to lines of sight, using bicubic B-splines. They applied their approach to the calibration of cameras but also of radiographic and range sensors.

Fiala and Basu used a 1D version of the two-plane model to calibrate the radial distortion profile of catadioptric cameras [143].

Lavest, Delherm et al. used the two-plane model to calibrate zoom cameras for different zoom settings and to obtain 3D reconstructions from images taken with different zoom factors but without moving the camera [109, 300, 301].

3.1.4 Models for Slit Cameras

Consider slit image acquisition proceeding by moving a 1D perspective camera with a constant translational velocity while acquiring images, which are stacked together to form a 2D image, i.e., a linear pushbroom panorama (cf. Section 2.1.1 and [197]). It is obviously an axial camera since all viewpoints lie on a straight line due to the camera's translational motion. In addition, all camera rays, for all acquisitions, cut the plane at infinity in the same line — the intersection of the 1D camera's view plane with the plane at infinity. Linear pushbroom cameras are thus a special case of two-slit camera, with one slit being a line at infinity [389].

Gupta and Hartley proposed models for the linear pushbroom panorama and its multi-view geometry [197]. They showed that forward projection can be modeled by a 3×4 projection matrix, which relates 3D point coordinates with lifted coordinates of the corresponding image point. They also showed expressions for back-projection. Let us re-derive them here. Let the 1D camera be modeled by the following back-projection operations, where $\mathbf{q}$ are the homogeneous coordinates of image points:

$$\begin{aligned}
\mathbf{B}^f &= \begin{pmatrix}\mathbf{t}\\1\end{pmatrix} + \frac{q_2}{q_3}\begin{pmatrix}\mathbf{v}\\0\end{pmatrix} \sim q_3\begin{pmatrix}\mathbf{t}\\1\end{pmatrix} + q_2\begin{pmatrix}\mathbf{v}\\0\end{pmatrix}\\
\mathbf{B}^i &= \begin{pmatrix}\mathbf{a} & \mathbf{0} & \mathbf{b}\end{pmatrix}\mathbf{q} = q_1\mathbf{a} + q_3\mathbf{b}.
\end{aligned}$$

Here, the second image coordinate is assumed to correspond to the acquisition time and $\mathbf{v}$ gives the translational velocity of the camera; $\mathbf{t}$ is the optical center position at the first acquisition and vectors $\mathbf{a}$ and $\mathbf{b}$ depend on the 1D camera's intrinsic parameters and orientation. The latter are constant, thus the point at infinity of a back-projected ray does not depend on the second image coordinate (which explains the column of zeroes in the second above equation).

The Plücker coordinates of camera rays may now be expressed as follows, based on the above two equations defining $\mathbf{B}^f$ and $\mathbf{B}^i$ (cf. Equation (1.2)):

$$\begin{aligned}
\mathbf{B}^l &= \begin{pmatrix}-q_3(q_1\mathbf{a} + q_3\mathbf{b})\\(q_3\mathbf{t} + q_2\mathbf{v}) \times (q_1\mathbf{a} + q_3\mathbf{b})\end{pmatrix}\\
&= \underbrace{\begin{pmatrix}\mathbf{0} & -\mathbf{a} & \mathbf{0} & -\mathbf{b}\\ \mathbf{v}\times\mathbf{a} & \mathbf{t}\times\mathbf{a} & \mathbf{v}\times\mathbf{b} & \mathbf{t}\times\mathbf{b}\end{pmatrix}}_{\mathsf{B}^l_{6\times 4}}\begin{pmatrix}q_1q_2\\q_1q_3\\q_2q_3\\q_3^2\end{pmatrix}.
\end{aligned}$$

The geometry of more general two-slit cameras was studied by Pajdla and Zomet et al. [389, 576]. Among other issues, Zomet et al. derived the forward projection equations and have shown that 3D lines are imaged as conics. Feldman et al. showed that the back-projection for the same two-slit cameras can be expressed via a 6×6 back-projection matrix, operating on second order lifted image coordinates [139]. As will

be seen later (Sections 4.2.1 and 4.3), once it is established that back-projection can be written as a linear mapping operating on order-2 lifted image coordinates, it follows automatically that 3D lines are imaged as conics and that there exists a fundamental matrix of size 6×6.

Models for other slit image systems are in principle straightforward to establish. Consider the case of a 1D camera rotated about some axis (not necessarily containing the 1D camera's optical center). Let the axis be the Z-axis and the optical center at the first acquisition ($q_2 = 0$) be at position $\mathbf{t}$. Further, let the intrinsic parameters and orientation of the 1D camera be subsumed by vectors $\mathbf{a}$ and $\mathbf{b}$, as above for the linear pushbroom case. Then, full back-projection operations can be written as:

$$\mathbf{B}^f = \begin{pmatrix} \cos(kq_2/q_3) & -\sin(kq_2/q_3) & 0 & 0 \\ \sin(kq_2/q_3) & \cos(kq_2/q_3) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ 1 \end{pmatrix}$$

$$\mathbf{B}^i = \begin{pmatrix} \cos(kq_2/q_3) & -\sin(kq_2/q_3) & 0 \\ \sin(kq_2/q_3) & \cos(kq_2/q_3) & 0 \\ 0 & 0 & 1 \end{pmatrix} (q_1\mathbf{a} + q_3\mathbf{b}).$$

Here, k expresses the rotational velocity and q_2/q_3 the time. Due to the trigonometric expressions in the rotation matrix, it is not possible to write the back-projection in polynomial form though and for example to establish fundamental matrices.

3.1.5 Models for Central Catadioptric Cameras

Geyer and Daniilidis proposed a unifying model for single-mirror central catadioptric systems (cf. Section 2.3.1) [164, 165]. It is embodied by a two-step projection procedure, defined by a unit sphere and a pinhole camera, as follows (cf. Figure 3.2). First, a 3D point is mapped onto the sphere, by intersecting the sphere with the line spanned by the point and the sphere's center. Second, that intersection point is projected into a (virtual) pinhole camera. Geyer and Daniilidis showed that this model subsumes all single-mirror central catadioptric systems

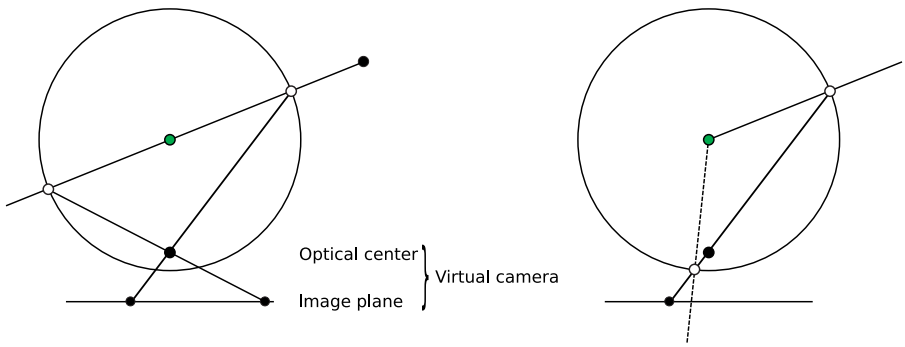


Fig. 3.2 Unified central catadioptric model of Geyer and Daniilidis. *Left*: Projection from 3D to 2D. The line joining the 3D point and the center of the sphere, cuts the sphere. The intersection points are then projected by a virtual pinhole camera. *Right*: Back-projection. An image point is first back-projected relative to the virtual pinhole camera. The intersections of the camera ray with the sphere are determined. The final camera rays are the lines spanned by the intersections points and the center of the sphere. For clarity, the second final camera ray is shown as a dashed line.

(see Section 2.3.1 for a list of these); the center of the sphere is the effective viewpoint of the modeled system. Note that the first step, projection onto the sphere, has two mathematical solutions; in practice, one usually knows which one to pick but when formulating multi-view geometry for instance, one may have to carry along both solutions. Also note that the virtual pinhole camera of the model is not equivalent to the camera looking at the actual mirror in the true catadioptric system; it is just an algebraic model.

Back-projection with this model can be formulated as follows. Without loss of generality, let us assume that the sphere is of radius 1 and centered in the origin. Let the projection matrix of the virtual pinhole camera be given as:

$$\mathsf{M}_{3\times 3}^{-1}\begin{pmatrix}\mathsf{I}_{3\times 3} & -\mathbf{t}\end{pmatrix}.$$

Then, an image point is back-projected, relative to the virtual pinhole, as:

$$\mathbf{B}_v^f \sim \begin{pmatrix}\mathbf{t}\\ 1\end{pmatrix}$$

$$\mathbf{B}_v^i \sim \mathsf{M}\mathbf{q}.$$

The back-projection ray intersects the sphere in two points, written in homogeneous coordinates as:

$$\begin{pmatrix} (\mathsf{M}\mathbf{q})^\mathsf{T}(\mathsf{M}\mathbf{q})\mathbf{t} - (\mathbf{t}^\mathsf{T}\mathsf{M}\mathbf{q})\mathsf{M}\mathbf{q} \pm \sqrt{D}\mathsf{M}\mathbf{q} \\ (\mathsf{M}\mathbf{q})^\mathsf{T}(\mathsf{M}\mathbf{q}) \end{pmatrix},$$

with $D = (\mathbf{t}^\mathsf{T}\mathsf{M}\mathbf{q})^2 - (\mathbf{t}^\mathsf{T}\mathbf{t} - 1)(\mathsf{M}\mathbf{q})(\mathsf{M}\mathbf{q})^\mathsf{T}$. The final back-projection directions are then given by the first three coordinates of these two points (since the sphere center is the origin):

$$\mathbf{B}^i \sim (\mathsf{M}\mathbf{q})^\mathsf{T}(\mathsf{M}\mathbf{q})\mathbf{t} - (\mathbf{t}^\mathsf{T}\mathsf{M}\mathbf{q})\mathsf{M}\mathbf{q} \pm \sqrt{D}\mathsf{M}\mathbf{q}.$$

Due to the square root in this expression, it is in general not possible to directly obtain polynomial expressions.[1] For the special case where the virtual pinhole lies on the sphere though, i.e., for $\mathbf{t}^\mathsf{T}\mathbf{t} = 1$, the two directions are:

$$\begin{aligned} \mathbf{B}_1^i &\sim \mathbf{t} \\ \mathbf{B}_2^i &\sim (\mathsf{M}\mathbf{q})^\mathsf{T}(\mathsf{M}\mathbf{q})\mathbf{t} - 2(\mathbf{t}^\mathsf{T}\mathsf{M}\mathbf{q})\mathsf{M}\mathbf{q}. \end{aligned}$$

This special case corresponds to para-catadioptric cameras. The first direction is independent of the image point $\mathbf{q}$ and can thus be discarded; the second one can be written via a 3×6 back-projection matrix B which depends on M and $\mathbf{t}$:

$$\mathbf{B}_2^i \sim \mathsf{B}_{3\times 6}^i \, \mathcal{L}^2(\mathbf{q}).$$

Let us finally consider the usual special case where the principal point is supposed to be known and to be the origin and the optical axis of the camera to be aligned with the mirror axis. Further, without loss of generality, we assume that the camera's optical center lies on the Z-axis. Then, $\mathsf{M} \sim \mathrm{diag}(1,1,k)$ and $\mathbf{t}^\mathsf{T} = (0,0,-1)$ and the back-projection becomes:

$$\mathbf{B}_2^i \sim \begin{pmatrix} 0 & 0 & 0 & -2k & 0 & 0 \\ 0 & 0 & 0 & 0 & -2k & 0 \\ 1 & 0 & 1 & 0 & 0 & -k^2 \end{pmatrix} \mathcal{L}^2(\mathbf{q}).$$

[1] However, Sturm and Barreto showed that polynomial back-projection expressions can be obtained by using degree-4 liftings of image coordinates [473].

The first and third columns are identical and the second one only contains zeroes. We can thus write the back-projection in a more compact form as:

$$\mathbf{B}_2^i \sim \begin{pmatrix} 0 & 2k & 0 & 0 \\ 0 & 0 & 2k & 0 \\ -1 & 0 & 0 & k^2 \end{pmatrix} \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}.$$

Geyer and Daniilidis analyzed various properties of catadioptric systems using their model, among which the projection of points and lines, the concepts of projective duality, epipolar geometry, etc. Barreto and Araújo proposed a modified version of this model and used it to study various geometric properties, especially those related to line images and their application for calibration [25, 28, 29]. Mei and Rives proposed to extend the model by adding classical terms for radial and tangential distortion [339].

A similar model was previously proposed by Smith et al. [453]. Their two-step model is defined by an arbitrary quadric surface, an effective viewpoint and a direction for an orthographic projection. A 3D point is first mapped onto the quadric surface by intersecting the line spanned by the point and the effective viewpoint, with the quadric. The intersection point is then orthographically projected to give the final image point. In practice, Smith et al. assumed that the quadric's highest point, in the direction of the orthogonal projection, lies above the effective viewpoint; the orthographic projection of the effective viewpoint itself is also considered as a distortion center. The final model used by Smith et al. has four parameters and it can model non-radially symmetric systems. It subsumes central para-catadioptric cameras as well as the division model (cf. Section 3.1.8) but not hyper-catadioptric cameras.

Ying and Hu proposed another extension by replacing the sphere with a general quadric of revolution [550]. One of the quadric's foci is chosen as effective viewpoint. The projection of a 3D point is modeled in two steps analogous to Geyer and Daniilidis' model: the line spanned by the 3D point and the effective viewpoint is cut by the quadric. The two intersection points are then projected into a virtual pinhole camera whose optical center is located on the mirror's revolution axis. This

gives the two theoretical images of a 3D point. Back-projection follows the inverse procedure. Ying and Hu showed that this model, besides central catadioptric cameras, also subsumes some fisheye models, such as stereographic projection (cf. Section 3.1.7) and the division model. A similar extension was proposed by Barreto, who showed that when replacing the sphere with a paraboloid, one may represent the division model [26, 27].

Bayro-Corrochano and López-Franco rewrote the model of Geyer and Daniilidis in terms of conformal geometry and gave the corresponding formulas for forward projection of points and lines and the back-projection of image points [44]. Tolvanen et al. expressed the model of Geyer and Daniilidis in terms of Clifford algebra [507]. Perwass and Sommer proposed a model similar to Geyer and Daniilidis', expressed in geometrical algebra terms; it subsumes the para-catadioptric and division models [399].

3.1.6 Non-central Catadioptric Cameras

Consider any non-central catadioptric system. Even if everything is known about the system, i.e., the shape of the mirror, the calibration of the camera looking at it as well as the relative camera–mirror pose, forward projection is a difficult problem. As for back-projection, it suffices to first back-project an image point relative to the camera, to find the (first) intersection point of the back-projected ray with the mirror and to reflect the ray according to Snell's law. For some setups, back-projection may be written analytically, e.g., for spherical or cone-shaped mirrors.

Forward projection is much harder since it is not immediately given at which point on the mirror the reflection happens. Again, for some cases, analytical models can be formulated but in general, there is no closed-form solution for this problem, so researchers have proposed to solve it by non-linear optimization. One possibility is for example to optimize the desired image point coordinates, where the cost function is the distance between the back-projected and reflected ray, and the 3D point whose projection is sought. Another possibility is to optimize the reflection point on the mirror, using for example a similar cost

function: the distance between the camera's optical center and the line joining that point and the 3D point, after reflecting that line in the mirror.

The special case of quadric-shaped mirrors was investigated by Gonçalves and Araújo who showed how to formulate the problem via three polynomials and to reduce it to an optimization in a single variable [179]. Vandeportaele also characterized the problem using polynomials and studied the number and nature of roots for the general as well as special cases of quadric-shaped mirrors [517]. Agrawal et al. contributed similar results for different types of quadric-shaped mirrors as well as for the case of cameras looking through refractive spheres [5].

3.1.7 Fisheye Models

Classical models. Several models were suggested for fisheye lenses (or rather, the opposite: fisheye lenses are constructed in order to satisfy such models) [48, 146, 226, 350, 536] — stereographic, equiangular,[2] sine-law, and equi-solid angle projection:

$$r_d = k\tan(\theta/2)$$
$$r_d = k\theta$$
$$r_d = k\sin\theta$$
$$r_d = k\sin(\theta/2).$$

Their back-projection equations are given in Table 3.1 (where the coefficient k is replaced by f to homogenize notations). We observe that stereographic back-projection can be written using polynomials; it can be written via a 3×4 back-projection matrix operating on lifted image coordinates (here, we suppose that the principal point is known and located at the origin) and that it is (not surprisingly) equivalent to the back-projection of para-catadioptric cameras. The other three models do not have back-projection expressions in terms of polynomials.

In [146], Fleck compared these models with respect to several criteria, among which the maximum theoretical field of view and light

[2] This is more often called equidistant model, but in the context of back-projection, we prefer to use the word equiangular.

fall-off. The stereographic model is preferred by Fleck, due to representing a good compromise between field of view, light fall-off, and a good approximate local preservation of Euclidean shapes. It seems though that most fisheyes are designed to approach the equiangular model, which is thus the model used most often for fisheye camera calibration.

Hall et al. for example used an equiangular model for a fisheye lens-based imaging system used for robot guidance [80, 382, 383].

An experimental evaluation of some of the above fisheye models for a particular camera was performed by Schneider et al. [439].

As a sidenote, let us remark that interestingly, some of the first scientific applications of fisheye lenses concerned the area of forest management (see `http://en.wikipedia.org/wiki/Hemispherical_photography`), where they were already used in the 1950s. Researchers in agricultural and forestry departments came up with calibration methods, independently from photogrammetrists. For example, Clark and Follin used the equiangular model for fisheyes [94]. Their application was forest management and panoramic images were used for assessing light interception and leaf areas in forests. Therefore, they were interested in images with equi-area properties, i.e., where the area of an image region is proportional to the solid angle of the corresponding field of view. In [94], they assessed how well different fisheye lenses approach this property.

Other models. Various other models were proposed for fisheye lenses, or lenses with a wide field of view in general, some of them aiming at subsuming several of the above classical models.

Herbert proposed the following alternative model for fisheye cameras [219]:

$$r_d = \theta + \sum_{i=1}^{5} a_i \sin \frac{i\,\pi\,\theta}{\theta_{\max}},$$

where $\theta_{\max}$ is the camera's field of view (assumed known).

Various fisheye lens designs and some performance characteristics were presented by Kumler and Bauer [294]. They used the following model:

$$r_d = k \sin(\beta\theta),$$

which encompasses the sine-law and equi-solid angle models shown above. Also, if β is very small, this model approaches the equiangular one.

A similar general model was proposed by Gennery [161]:

$$r_d = \frac{\sin(\beta\theta)}{k\cos(\max(0, \beta\theta))}.$$

For $\beta = 1$ this gives the pinhole model, for $\beta = 0.5$ stereographic projection, $\beta = -0.5$ gives equi-solid angle projection, and $\beta = -1$ the sine-law projection. As β approaches zero, this model approaches the equiangular one. Gennery suggested to couple this model (which subsumes ideal fisheye models) with a classical polynomial distortion model. Gennery's full model includes also the position of the effective optical center along the optical axis, as a function of θ or r_d; his full model is thus a non-central one (an axial model to be precise) and resembles in that respect that of Tardif et al. [491], see Section 3.3.2. Besides this effect, Gennery also included the orientation of the optical axis relative to the image plane, in order to model non-perpendicularity.

Yet another similar model was suggested by Bakstein and Pajdla [23], who used a combination of the stereographic and equi-solid angle models, depending on four parameters:

$$r_d = a\tan\frac{\theta}{b} + c\sin\frac{\theta}{d}.$$

Basu and Licardie proposed a distortion function called fisheye transform (FET) [41], inspired by the work of Schwartz on the geometry of the human visual system [441]:

$$\begin{aligned} r_d &= s\, log(1 + \lambda r_u) \\ r_u &= \frac{e^{r_d/s} - 1}{\lambda}. \end{aligned}$$

Although it is called *fisheye* transform, this model does not handle fields of view of 180° or higher.

Devernay and Faugeras proposed the FOV model (field-of-view model) [113]:

$$r_u = \frac{\tan(r_d\omega)}{2\tan\frac{\omega}{2}}.$$

According to their description, the model is equivalent to the equiangular model. However, this seems to be the case only for particular values of the parameter ω (cf. Figure 3.4(c) in Section 3.5). If we include the focal length of the perspective projection in the model, we get the following expression for the back-projection angle θ:

$$\theta = \operatorname{atan} \frac{\tan(r_d \omega)}{2f \tan \frac{\omega}{2}}.$$

For

$$f = \frac{1}{2 \tan \frac{1}{2\omega}}$$

the FOV model becomes equivalent to the equiangular model:

$$\theta = r_d \omega.$$

Like Gennery (see above), Devernay and Faugeras suggested to complement their model (or other fisheye models) by additional terms, e.g., the classical polynomial radial distortion model, to calibrate cameras.

3.1.8 Polynomial and Rational Polynomial Models

Most of the models in the previous section were based on trigonometric functions, whereas polynomial expressions were used in the classical distortion models of Section 3.1.1. The two-plane models of Section 3.1.3 were also embodied by polynomials. In this section, we review several other polynomial-based models.

A first possibility is to model radial distortion using polynomials, but instead of relating the distorted and undistorted radial distances as in Equation (3.3), one considers the relation between distorted radial distances and incidence angles between camera rays and the optical axis:

$$r_d = \sum_{i=1}^{n} c_i \theta^i. \tag{3.7}$$

This model was for example used by Herbert [219], Xiong and Turkowksi [539], Kannala and Brandt [266], Kannada et al. [267], Ying et al. [553], and Stehle et al. [459]. Kannala and Brandt further added

non-polynomial terms to account for tangential distortions [266]. They also showed an approximation of the inverse (backward) model. Note that the model in Equation (3.7) subsumes the equiangular model, which corresponds to using only the first coefficient c_1.

The reciprocal is of course also possible, i.e., models of the form:

$$\theta = \sum_{i=1}^{n} c_i \, r_d^i \ .$$

This was used for example by Inoue et al. where a 12-th degree polynomial was found to be suitable to calibrate a fisheye converter [245].

Beauchemin et al. proposed a polynomial model, but where polynomials relate polar image point coordinates before and after distortion, instead of Cartesian ones [46, 47]. Their model consists of two polynomials, one for the radial distortion function (distance from distortion center), one for the radial angle, capturing tangential distortions.

Hellmeier as well as Remy et al. used splines to model the 1D distortion function [218, 427].

Rational polynomials for radial distortion. Finally, several researchers have proposed models based on rational polynomials, as follows.

Fitzgibbon and independently, Bräuer-Burchardt and Voss, proposed the so-called division model, which has one distortion coefficient c [145, 62]. Before this, that model was already used by Lenz and Tsai [305, 306] (see also the textbook [253, 254] by Jähne); however, they used it as forward model, i.e., in the below equations, one would swap the forward and back-projection equations. The division model as defined by Fitzgibbon and Bräuer-Burchardt and Voss, is given as:

$$r_u = \frac{r_d}{1 - c r_d^2}$$

$$r_d = \frac{\sqrt{1 + 4 c r_u^2} - 1}{2 c r_u}.$$

A nice feature is that the model has a closed-form inverse.

The division model can be generalized by using general rational polynomials to relate r_u and r_d, as proposed by Ma et al. [327]. Rational

polynomials were already used before the introduction of the division model, e.g., by Hellmeier [218].

Mičušik and Pajdla [346] and Mičušik [344] used a similar model for the relation between image distance r_d and the angle between optical axis and camera rays, instead of the relation between image distances in distorted and undistorted images, as above. Their model has two parameters and is of the form:

$$\theta = \frac{a r_d}{1 + b r_d^2}.$$

It is in some sense a division model for the equiangular fisheye model, which it extends. Like for the equiangular model, expressing back-projection directions involves trigonometric functions; hence, Mičušik and Pajdla proposed to linearize their back-projection model using a Taylor development in order to perform an initial estimation of the model parameters during (self-)calibration.

Rational polynomials for full back-projection models. Instead of using rational functions to model the 1D distortion function, as above, one may use them for the actual back-projection mapping, as foreshadowed by the two-plane methods of Chen and Martins et al. [89, 90, 331] (see Section 3.1.3) and fully introduced by Claus and Fitzgibbon [98]. In general, one may write general mappings between distorted and undistorted image points:

$$\mathcal{L}^a(\mathbf{q_u}) \sim \mathsf{M}\, \mathcal{L}^b(\mathbf{q_d}),$$

where the matrix M would contain the coefficients of homogeneous polynomials relating the lifted coordinates in the distorted and undistorted image planes (for full generality, we assume that different liftings may be possible, i.e., $a \neq b$). We speak of *rational* polynomial models here since "undoing" homogeneous point coordinates involves divisions of coordinates.

It is more common though to use this kind of model directly for the back-projection mapping of distorted image points to back-projection directions or rays:

$$\mathbf{B}^i \sim \mathsf{B}^i\, \mathcal{L}^b(\mathbf{q_d}), \tag{3.8}$$

$$\mathbf{B}^l \sim \mathsf{B}^l\, \mathcal{L}^b(\mathbf{q_d}). \tag{3.9}$$

In practice, Claus and Fitzgibbon used a bi-quadratic such model for central cameras, i.e., $b = 2$ and B^i of size 3×6 [98]. It is easy to see that this encompasses the division model as well as the para-catadioptric and stereographic models. Note that in order to represent the classical radial distortion model, higher order liftings are required (cubic in the case of a single distortion coefficient, cf. Section 3.1.1).

Let us also note that the two-plane models of Chen and Martins et al. (see Section 3.1.3) are instances of rational polynomial back-projection models.

It may be worthwhile to study which instances of the models (Equations (3.8) and (3.9)) correspond for example to radially symmetric, central, perspective, or otherwise special camera models.

Rational polynomial forward models. Rational polynomial models are usually either applied for the back-projection function as proposed by Claus and Fitzgibbon [98], or for the 1D radial undistortion function as proposed by Ma et al. [327]. Hartley and Saxena also proposed a rational camera model, but for the forward projection function [210]: the image point coordinates are expressed via rational polynomial functions of the 3D point coordinates (Hartley and Saxena used cubic polynomials). In the general case, this corresponds to curved camera rays, i.e., the set of 3D points that get projected onto the same image point, form a curve. Hartley and Saxena showed that this model is a good fit for SAR cameras. It also subsumes "linear" cameras such as perspective and linear pushbroom ones, although fitting it to data obtained from these may lead to overfitting due to a strong overparameterization.

Essentially the same approach seems to be in routine use to model satellite imagery, cf. [147]. In Section 11.3 of that reference, the general term of "replacement sensor models" is employed to denote any camera model that replaces a rigorous, physics-based, model by empirical expressions. The most used ones are rational polynomial functions, much like in the above works.

Hall et al. proposed the RTcam (rational tensor camera model) [200]. It essentially proceeds by first transforming 3D points via bi-quadratic rational functions, followed by a perspective projection to a

2D image plane (the full RTcam model is defined for arbitrary dimensions of scene and image). Together, the two imaging steps still correspond to bi-quadratic rational functions. Similar to the GLC model (see Section 3.4.4), the development of the RTcam is mainly motivated by applications in stylized, or non-photorealistic, rendering.

3.1.9 One-Dimensional Radial Models

Thirthala and Pollefeys formally introduced so-called 1D radial camera models [500, 501]. They hinge on the existence of an optical axis and a center of distortion. Planes containing the optical axis are called radial planes and lines going through the distortion center, radial lines. The only main assumption about the projection function is that all 3D points in a radial plane are imaged to points on a same radial line. This effectively allows completely arbitrary radial distortions, and especially, does not require any radial symmetry of the distortion (the distortion "function" may be different for each radial plane–line pair) and it allows for non-central projections.

To make the model tractable, a second assumption is used though, namely that the projection can be modeled by a 2×4 projection matrix, mapping 3D points onto radial lines, represented by two homogeneous coordinates. In other words, it is assumed that there exists a projective relationship between the pencil of radial planes in 3D and that of radial lines in the image. Thirthala and Pollefeys showed that quadrifocal tensors can be formulated based on these projection matrices, and used for self-calibrating cameras. In general, no matching constraints exist for less than four images. However, for the case of pure rotational camera motion, a trifocal tensor exists and can also be used for self-calibration.

One-dimensional radial models have actually been used somewhat implicitly before, especially embodied within the radial alignment constraint (RAC) of Tsai [515], used also in other works [468, 106]. Tsai proposed a camera calibration approach for cameras with radial distortion, where one of the initial steps concerns the estimation of the calibration grid's pose relative to the camera. At this stage, it is assumed that the camera follows a 1D radial model and that the

distortion center is known. Compared to the general model of Thirthala and Pollefeys, Tsai further assumed that angles between radial lines and associated radial planes are identical (e.g., he had to know the aspect ratio). Using these assumptions, the radial plane associated with each extracted calibration point can be computed. One may then estimate the pose of the calibration grid using the constraint that with the correct pose, each point on the grid, must lie in the radial plane of the matched image point. It is clear that the pose cannot be fully computed: any translation along the optical axis will still satisfy the above constraint. Hence, 5 degrees of freedom are computed, from a minimum of 5 point matches.

A 1D radial model assumption was also used in the calibration approach of Hartley and Kang [209]. In the initial step, they used this to compute the distortion center from the image of a calibration grid; see more on their approach in Section 5.1.2.

3.1.10 Neural Networks

The projection or back-projection function has also been modeled and computed using neural networks, see for example [9, 7, 535, 577]. This is an interesting approach since it is inherently flexible and can be applied to calibrate various kinds of cameras with various types of distortion.

3.2 Local Camera Models

3.2.1 Two-Plane Model

The third version of the two-plane model proposed by Chen and Martins et al. [89, 90, 331] (cf. Section 3.1.3) is a local camera model. Here, for each of the two calibration images, the extracted matches define triangulations of the image and calibration grids respectively, cf. Figure 3.3. Back-projection is now modeled separately for each pair of matching triangles, by an affine transformation as in the first version of the two-plane model, computed from the matches associated with the three triangle vertices. To back-project an image point, the following is done. For each calibration image, the triangle into which the image point falls is determined. A point on the calibration grid is computed

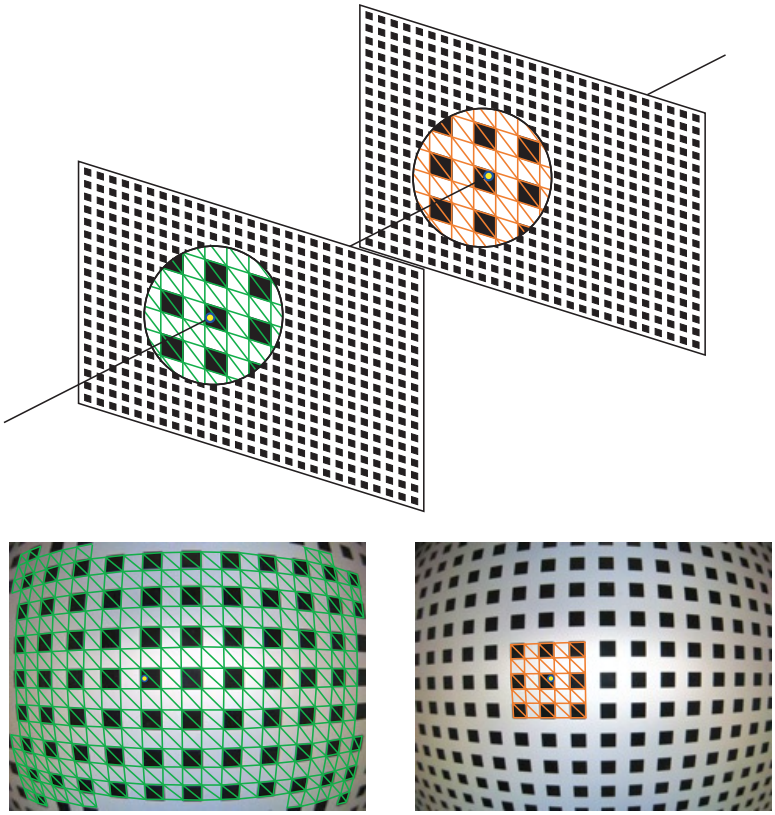


Fig. 3.3 Illustration of a local two-plane model (cf. text). For the second image, the triangulation is shown only partially.

via the affine transformation associated with that triangle. This is done separately for both calibration images/grids. The back-projected camera ray is then computed by spanning the line between the points on the two calibration grids.

3.2.2 Warping-Based Models

Very similar to the two-plane approach is the following idea. Whereas the two-plane model interpolates 3D points on calibration grids, the following methods perform an interpolation in the image domain. They can all be applied with a single image of a grid, as for the correction of non-perspective distortions.

The general idea is as follows: consider one calibration plane and an image thereof, as well as matches between the plane and the image. Consider any warping scheme that maps the acquired image into the "ideal" image of the calibration plane (let's say, a texture map of the plane). Since the ideal image is distortion-free (it is a perspective image), this warping thus removes non-perspective distortions. If we determine the parameters of this warping using the acquired image, we may then apply it to any other image and thus remove its non-perspective distortions. This gives images which are distortion-corrected, i.e., images as acquired by some perspective camera. However, we don't yet know the associated perspective camera. Still, when acquiring and then warping one or more images, the resulting images correspond to images acquired by the same perspective camera. This enables the application of any algorithm valid for perspective cameras, to these images. For example, if images of a calibration grid are acquired (be it planar or three-dimensional), any classical perspective calibration algorithm may be applied to calibrate the virtual camera. The same holds true for self-calibration. The complete calibration of the true camera is then the combination of the warp, with the computed perspective (self-)calibration.

Several methods following this general idea were proposed; they mostly differ in the choice of interpolation scheme which is required to warp entire images (the role played by the piecewise affine warp in the two-plane method). Without going into details, we cite the approaches of Green et al. [186], Goshtasby [182], Yokobori et al. [557], Brand et al. [59, 60], and Jackowski et al. [252]. Peuchot and Saint-André [401] and Bogner [54] also mentioned the feasibility of this type of approach. The method of Jackowski et al. corrects for color non-linearities in addition to geometrical distortions [252].

Sagawa et al. used the same idea and combined it with a structured light-type approach in order to get dense matches between the image and the calibration grid (a flat screen) [433], an approach used also for example by Grossberg and Nayar, Tardif et al. and Dunne et al. (see Section 3.3). From dense matches, the distortion-corrected image can be constructed pixel-by-pixel instead of requiring piecewise warps and an interpolation scheme.

It seems somewhat surprising that apparently, none of the above works went through the last obvious step to obtain a complete camera calibration. Namely, given these distortion-corrected, hence "virtual" perspective images, one may apply any calibration approach for pinhole cameras to calibrate the perspective part of the projection in addition to the non-perspective one. Such approaches were proposed previously, but only in combination with a plumb-line type distortion correction based on classical polynomial distortion models, see e.g., the work of Grosky and Tamburino [190]. Dunne et al. seem to be the first to combine the above generic distortion correction with a subsequent pinhole calibration [125]. They also used a structured-light type approach to get dense, per-pixel, matches between calibration grids and images; due to the per-pixel nature of the approach, we consider it in the discrete class of models and describe it in the next section.

Comment. All approaches in this section are valid for central cameras only since they are based on the idea of (bijective) warps of distorted images into perspective ones. The two-plane approach of Section 3.2.1 is in some sense a non-central counterpart.

3.2.3 Other Local Models

Qiu and Ma proposed an approach that is similar in spirit to the warp-based ones described above [410]. They used as input an image of a 3D calibration object, as opposed to a planar grid as in most other similar approaches. From 2D-to-3D matches, they then seemed to compute a best-fit perspective projection matrix for that image and projected the 3D calibration points using it. The result will constitute the ideal image onto which the distorted input image will be warped, for distortion correction. Input to the next step of calibration are the images of calibration points in the distorted image and in the ideal image, respectively. The goal is then to compute distortion functions Δx and Δy such that:

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \begin{pmatrix} x_u + \Delta x(x_u, y_u) \\ y_u + \Delta y(x_u, y_u) \end{pmatrix}.$$

In [410], the distortion functions were estimated at a regular lattice of image pixels using non-parametric regression, and extended to the whole image by linear or nearest-neighbor interpolation.

As mentioned above, Qiu and Ma used a 3D calibration object and fitted a perspective projection matrix to the 2D-to-3D matches. The approach could easily be adopted to using a 2D calibration grid: then one may directly use the model of the grid as ideal image, such as in most other similar approaches or, one may compute the best-fit homography (instead of projection matrix) between the 2D-to-2D matches and use it to project the calibration points into an ideal image that is "closer" to the actual input image. The rest of the approach would be strictly identical.

Ragot proposed a similar idea, where 3D-to-2D matches obtained from an image of a calibration grid are interpolated to determine the forward respectively back-projection of generic 3D respectively image points [412]. The interpolation is done within triangular tessellations of 3D respectively image points.

Munjy proposed another very similar approach, based on the finite element method [359, 360]. He considered a regular triangular or rectangular tessellation of the image plane and modeled the distortion function with the help of one focal length per vertex (giving the angle between the line of sight associated with the vertex, and the optical axis). Focal lengths of all other image points can then be computed via bilinear interpolation of the focal lengths of the vertices of the triangle/rectangle in which image points lie. The parameters of the model (principal point, focal lengths of vertices) are computed by bundle adjustment (can be done when using a calibration grid but also in full self-calibration mode where 3D point coordinates are also optimized, using multiple images).

Lichti and Chapman extended this approach by coupling it with the classical radial and tangential distortion terms [315]. Back-projection is thus carried out by correcting for these classical distortions and then back-projecting the resulting points using the local focal length computed from the finite-element model. Lichti and Chapman also discussed correlations among the model's parameters and how to acquire images in order to minimize them. In [316], they further extended the

approach by including continuity constraints in order to overcome discontinuities of the distortion function due to the bilinear interpolation carried out over neighboring rectangles.

Schreier et al. [440] and Cornille [101] used a similar model, where B-splines are used to map positions in distorted images to those in undistorted ones. They used that approach to calibrate optical and scanning electron microscopes.

Stevenson and Fleck computed a 2D distortion function that is piecewise over a triangular tessellation of the image plane [463]. The basic idea is to apply an affine warp to each individual triangle such that the image becomes distortion-free. A notable difference with most other similar works is that the distortion-free image is generated with respect to stereographic projection instead of perspective, i.e., distortion-free does not mean as usual that straight lines remain straight (as in perspective projection) but rather that any sphere is imaged as a circle (a property of stereographic projection). Consequently, the input are images of spherical objects. In a first step, ellipses are fitted to their outlines in the images (which is an approximation, precluding the use of too large spheres). Then, the Delaunay triangulation of the centers of all ellipses is computed. Let us now consider one triangle and the associated three ellipses, centered in the triangle's vertices. We may compute an affine transformation that maps the three ellipses as closely as possible into circles. This transformation is naturally only defined up to a scale change ("circle-ness" does not depend on scale). The goal is then to find a new triangulation that is consistent with the original one, *modulo* these local affine transformations. This new triangulation would be nothing else than the stereographic image of the set of sphere centers. This description does not exactly correspond to the approach taken by Stevenson and Fleck, but transcribes its basic idea. Once the new triangulation computed, the input image can be warped into a stereographic image by piecewise warping it from the original into the new triangulation. The result is an uncalibrated stereographic image, i.e., whose "focal length" (overall scale) is arbitrary and whose distortion center is unknown.

The approach is somewhat similar in spirit to the warping-based ones of Section 3.2.2, in that it results in a piecewise warp of an image,

allowing to undistort it. Differences however are the use of stereographic projection as reference instead of perspective and that the warp is not computed on the basis of an ideal input image (the model of a 2D calibration grid), but using images of 3D primitives (here, spheres).

3.3 Discrete Camera Models

3.3.1 Ray-Based Models

As explained in the introduction to this section, most discrete camera models can be seen as the limiting case of the above local models, where interpolation is required in the immediate vicinity of pixels instead of in larger image regions. The main difference between the following discrete approaches and the local ones is the explicit motivation to calibrate individual camera rays, as opposed to forward or back-projection mappings.

The idea of calibrating individual camera rays for discrete image locations was formulated by several researchers. Among the first were probably Gremban et al. who proposed extensions of the two-plane method in [189] and discussed the possibility of a ray-based calibration. However, they concluded that "A lookup table of calibration data for each pixel would be prohibitively expensive". This has to be understood in the historical context; such a lookup table requires four coefficients per pixel, i.e., is of a size which does not pose memory problems anymore nowadays. As alternative solution, Gremban et al. effectively proposed global and local camera models, based on global and local interpolation schemes to compute camera rays for any pixel: global schemes are based on a function that is valid for the whole image plane whereas local schemes look up the closest image points from calibration data (in the case of back-projection) or the closest lines of sight (in the case of forward projection) and interpolate accordingly.

Using structured light-type acquisition setups, such interpolations can be avoided altogether, by matching image points directly to calibration points. This has been used by various researchers, e.g., Grossberg and Nayar [191], Dunne et al. [123, 124, 125], Sagawa et al. [433] (cf. Section 3.2.2), and Tardif et al. [491].

Southwell et al. proposed an idea for a structured-light type ray-based camera calibration approach [455]. The idea was to insert the camera in a cylinder painted on the inside in smoothly varying color shades such that each "point" on the cylinder has a unique color. Then, for each camera pixel, the scene point it sees can in theory be uniquely determined. Southwell et al. proposed to calibrate a camera by generating a look-up table where for each pixel, the direction of its camera ray is stored, after computing it from the matched point on the cylinder. They did not use this approach in practice though, due to practical and signal processing problems.

Grossberg and Nayar proposed a generic imaging model consisting of geometric as well as radiometric and optical parts [191]. The geometric part is identical in spirit to the two-plane model and its successors. However, Grossberg and Nayar calibrated pixels individually, without any interpolation, using structured light-type approaches allowing to densely match images and calibration objets (e.g., flat screens). Further, instead of computing lines of sight for individual pixels, they actually computed half-lines. This was mainly achieved by computing and using the caustic of the imaging system which is used as the hull of the imaging system's model, from which camera rays emanate in an outward direction. Besides the geometric part of the imaging model, it also contains radiometric and optical aspects: for example, each pixel is associated with individual radiometric response and point spread functions. Overall, a line of sight (or rather, a half-line), together with these non-geometric properties, make what Grossberg and Nayar termed a *raxel*, a sort of tiny camera associated with each individual pixel of the imaging system.

Such a ray-based imaging model has been used recently by several researchers to devise calibration and structure-from-motion algorithms, for example by Ramalingam et al. [414, 415, 418, 477], Dunne et al. [123, 124, 125], and Gonçalves and Araújo [180]. Ray-based calibration approaches are described in Section 5.1.2.

Debaecker et al. proposed to model the field of view of each pixel by a cone instead of a (half-)line in order to model the fact that pixels gather a volume of light rays [108]. This may be seen as an alternative to using a point spread function like Grossberg and Nayar.

Comments. In the above works, it is usually proposed to compute one camera ray per image pixel. Note however that this choice, while being intuitive, is not the only one for a discrete imaging model. Indeed, one may compute rays for any other discrete sample of image locations. With a regular super-pixel-based sampling for example, one ray for every n-th pixel would be computed; to compute rays for every pixel, an interpolation scheme would then be required, i.e., we are looking at local camera models, see the previous section. Even with a per-pixel sampling, interpolation is required to perform back-projection for every image *point*. Finally, nothing prevents in principle to apply a sub-pixel sampling for ray-based calibration. The appropriate choice of sampling density is a model selection question and depends on several factors, see the general discussion in Section 3.6.

3.3.2 Discrete Sampling of the Distortion Curve

Instead of computing camera rays at a discrete set of image locations, which is nothing else than gathering samples of a generic back-projection mapping, one may also gather such samples for any variant of the 1D (un) distortion function: $r_d(\theta), \theta(r_d), r_d(r_u)$, or $r_u(r_d)$.

Most calibration approaches based on this idea then fit some global distortion model to these samples. Examples are some collimator-based calibration approaches (cf. Section 5.1.4) or an early approach for plane-based calibration by Hallert (cf. Section 5.1.2) [201].

More recent such approaches are the ones by Stevenson and Fleck [462] (see Section 5.3.2), Hartley and Kang [209] and Tardif et al. [491] (see Section 5.1.2), as well as Ying and Hu [552] (see Section 5.1.1). In addition to modeling the usual 1D distortion function, Tardif et al. and Ying and Hu also considered a non-central radial distortion model: to each radial distance r_d, is associated an angle θ as well as an optical center on the optical axis. This is thus a radially symmetric axial model, similar to the one proposed by Gennery (cf. Section 3.1.7) [161].

3.3.3 "Isolated" Camera Rays

Grossmann et al. have considered what might be the most discrete camera model, consisting of a sparse set of pixels and associated camera

rays, with no requirement that rays have close-by neighboring rays [192, 194]. Their proposed self-calibration approach (see Section 5.3.3) only makes the assumption that the camera is central.

3.4 Models for the Distribution of Camera Rays

What a camera sees of a scene depends on the light rays it captures, not on its intrinsic geometry, i.e., which pixel exactly catches the photons of which light ray. In that respect, we may split the camera's intrinsic parameters into two sets: one that explains which light rays the camera captures (these will then be considered as camera rays) and one that explains the mapping between camera rays and image points. To make the distinction clearer, let us consider any two different central cameras, e.g., a perspective camera and a central catadioptric one. Being central, both cameras catch exactly the same set of light rays: all light rays that contain the optical center (here, we neglect of course the finite aperture of our cameras and their different fields of view). What differs is the mapping from image locations to rays and consequences thereof (e.g., non-perspective distortions and the spatial/angular resolution profile). On top of this come the extrinsic parameters, telling the global orientation and position of the camera, or in other words, of its set of camera rays.

In this section, we summarize several works which are primarily concerned with the first part of the "intrinsic geometry" of cameras, namely a description of which rays a camera captures: the locus, or spatial distribution, of camera rays. Several taxonomies and computational models have been proposed in the last years [390, 405, 445, 472, 559]. The main motivation of these works is to reason on the spatial distribution of camera rays; the mapping between rays and image points is of secondary importance although some of these works also propose such a mapping in order to obtain a full camera model.

3.4.1 From Central to Oblique Cameras

Cameras may be classified according to the existence of a geometric entity that is incident with all camera rays [390, 405, 472]. For a **central** camera, there exists a 3D point, the optical center, that is incident

with all rays. Any other camera is **non-central**. Among non-central cameras, several subgroups exist, as follows.

In an **axial** camera, all rays cut some real 3D line. A special case is the **two-slit** camera, where there exist two real lines that cut all rays (for example, linear pushbroom panoramas, cf. Section 2.1.1). Cameras falling outside any of these categories are usually termed **fully non-central** or the like. A special case here are **oblique** cameras, for which any two camera rays are either identical or skew [390]. Oblique cameras are the most non-central cameras possible. A special case are **linear oblique** cameras, see Section 3.4.3, for which all rays are cut by an imaginary 3D line [389] (in fact, by two conjugated imaginary lines [405]).

Other cases are cameras, whose rays are incident with a circle, as is the case with circular non-central mosaics (cf. Section 2.1.1). Another example is catadioptric camera consisting of a cone-shaped mirror and a camera located on the mirror axis. All reflected rays are incident with a circle centered in the mirror axis. However, such systems are also of the axial type, since all reflected rays also intersect the mirror axis.

One may define many other categories of cameras, by choosing other geometric primitives to which all rays are incident.

3.4.2 Ray Distributions Allowing for Standard Stereo Geometry

Pajdla [390] and Seitz and Kim [445] were interested in studying which sets of camera rays admit stereo imaging conditions. What is meant by this? To explain this, we look ahead to Section 4 on epipolar geometry. The basic question of epipolar geometry and stereovision is, given a point in one image, which are the potentially matching points in the other image. The answer to this is: all points whose camera rays intersect the ray of the first point. These define an epipolar curve in the second image. Epipolar curves in the first image are defined analogously. Essentially, two cameras are in standard stereo geometry if epipolar curves in the two images are in a one-to-one relationship: the camera ray of any point on an epipolar curve in one image cuts all camera rays of points on the related epipolar curve in the other image

(for a more precise definition, see [445]). An auxiliary requirement is that the set of epipolar curves in each image is one dimensional.

Not surprisingly perhaps, the set of rays of two cameras in standard stereo geometry, lie on a set of doubly ruled quadrics[3] [390, 445]. The two individual cameras then correspond to the two rulings of these quadrics. Note that this comprises several types of non-central cameras, such as circular non-central mosaics and others [445] as well as linear oblique cameras [390].

This result concerns the locus of camera rays and is independent of the mapping from rays to image points. What it implies is that for any two cameras in standard stereo geometry, the acquired images may be warped into images where epipolar curves are horizontal lines and matching epipolar lines have the same vertical coordinate. In other words, it is possible to rectify images and then apply any standard stereo algorithm.

3.4.3 Linear Oblique Cameras

Pajdla studied the case of *linear oblique cameras*, already mentioned above [390]. A defining characteristic of these is the existence of a linear transformation that maps each 3D point to the unique camera ray containing that point. Pajdla showed that all linear oblique cameras are projectively equivalent to the one defined as follows: the camera ray containing a 3D point $(X, Y, Z, W)^\mathsf{T}$ is spanned by that point and the point with coordinates $(Y, -X, W, -Z)^\mathsf{T}$.

This defines the set of camera rays in a linear oblique camera. In order to get a full camera model, including the mapping between rays and image points, Pajdla proposed to use a plane as reference. Let $\mathbf{q}$ be the homogeneous 2D coordinates of a point in the reference plane (the image plane if one likes). Each point inside the plane is mapped to a 3D point by some 4×3 pose matrix M:

$$\mathbf{Q} \sim \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \sim \mathsf{M}\mathbf{q}.$$

[3] For central cameras, this is the pencil of epipolar planes.

The camera ray is then spanned by $\mathbf{Q}$ and the second 3D point, defined above. It is easy to prove that the camera ray can be written as:

$$\mathbf{B}^l \sim \mathsf{B}^l \mathcal{L}^2(\mathbf{q}),$$

with B^l a 6×6 back-projection matrix depending on M, operating on lifted image point coordinates. The back-projection matrix is always of rank at most 4.

For example, if we choose the following pose matrix:

$$\mathsf{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which corresponds to choosing the XY-plane as reference plane, then one obtains the following back-projection matrix:

$$\mathsf{B}^l \sim \begin{pmatrix} 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.10}$$

We may see that, similarly to what was shown for the classical radial distortion model (cf. Section 3.1.1) or para-catadioptric cameras (cf. Section 3.1.5), the back-projection matrix may be compacted. The second column only contains zeroes and the first and third ones are identical. Hence, back-projection for linear oblique cameras may be written as:

$$\mathbf{B}^l \sim \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}.$$

Observation. Linear oblique cameras are actually subsumed by the linear two-plane model of Chen and Martins et al. (cf. Section 3.1.3): it is easy to prove that the back-projection matrix of Equation (3.10) can be obtained from the one of the two-plane model (cf. Equation (3.6)) by an appropriate choice of affine transformations A_1 and A_2.

3.4.4 General Linear Camera Model (GLC)

Yu and McMillan proposed the so-called General Linear Camera model (GLC) [559, 558], which parameterizes the set of camera rays by affine transformations of three basis rays. Concretely, they adopted a two-plane parameterization of rays (see also a paper by Gu et al. [196] on this parameterization) as follows. The three basis rays are given by points:

$$\mathbf{A}_i = \begin{pmatrix} u_i \\ v_i \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{B}_i = \begin{pmatrix} s_i \\ t_i \\ 1 \\ 1 \end{pmatrix},$$

where we assume that the two planes are $Z = 0$ and $Z = 1$. All other rays are then given by two parameters α and β, defining points in the two planes via affine transformations of the basis points parameterized by two scalars α and β:

$$\mathbf{A} = \begin{pmatrix} \alpha u_1 + \beta u_2 + (1-\alpha-\beta)u_3 \\ \alpha v_1 + \beta v_2 + (1-\alpha-\beta)v_3 \\ 0 \\ 1 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} \alpha s_1 + \beta s_2 + (1-\alpha-\beta)s_3 \\ \alpha t_1 + \beta t_2 + (1-\alpha-\beta)t_3 \\ 1 \\ 1 \end{pmatrix}.$$

Yu and McMillan showed that this model allows to generate the set of camera rays of several imaging geometries, such as pinhole and orthographic cameras but also non-central cameras such as some two-slit ones. It also allows to generate less classical imaging geometries, called by Yu and McMillan the pencil camera, the twisted orthographic camera, and the bilinear camera.

Like for linear oblique cameras, the initial main motivation of the GLC model is to propose a model for the locus of camera rays, not necessarily a full camera model. For linear oblique cameras, a way to

get such a full model, mapping image points to rays, was shown above. As for the GLC, a natural choice is to consider the two parameters α and β as 2D image coordinates. Before going further, let us replace the image coordinates by homogeneous ones such that:

$$\mathbf{q} \sim \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}.$$

Then:

$$\mathbf{A} \sim \begin{pmatrix} (u_1 - u_3)q_1 + (u_2 - u_3)q_2 + u_3 q_3 \\ (v_1 - v_3)q_1 + (v_2 - v_3)q_2 + v_3 q_3 \\ 0 \\ q_3 \end{pmatrix} \quad \mathbf{B} \sim \begin{pmatrix} (s_1 - s_3)q_1 + (s_2 - s_3)q_2 + s_3 q_3 \\ (t_1 - t_3)q_1 + (t_2 - t_3)q_2 + t_3 q_3 \\ q_3 \\ q_3 \end{pmatrix}.$$

Finally, the camera ray, spanned by $\mathbf{A}$ and $\mathbf{B}$, can be written via the following 6×6 back-projection matrix, operating on lifted image coordinates $\mathcal{L}^2(\mathbf{q})$:

$$\mathsf{B}^l = \begin{pmatrix} 0 & 0 & 0 & u_1 - u_3 - s_1 + s_3 & u_2 - u_3 - s_2 + s_3 & u_3 - s_3 \\ 0 & 0 & 0 & v_1 - v_3 - t_1 + t_3 & v_2 - v_3 - t_2 + t_3 & v_3 - t_3 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & v_1 - v_3 & v_2 - v_3 & v_3 \\ 0 & 0 & 0 & u_3 - u_1 & u_3 - u_2 & -u_3 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \end{pmatrix}. \tag{3.11}$$

For the sake of conciseness, the coefficients b_i are not shown here; they can be easily computed if necessary.

Observations. We see that the rays in a GLC can be modeled by a quadratic rational camera model, cf. Section 3.1.8. Furthermore, by comparing the back-projection matrices of the GLC and the linear version of the two-plane model (Equation (3.6) in Section 3.1.3), we observe that they have exactly the same form. It can indeed be proven that the two models are equivalent. This is maybe not surprising since both models are based on affine transformations in two reference planes, although they are parameterized in different ways. Whereas the motivation for developing the two-plane model of Chen and Martins et al.

was to interpolate between calibration data, Yu and McMillan were aiming at developing novel camera geometries. Compared to works on the two-plane model, Yu and McMillan carried out an in-depth study of the modeled camera geometries.

Other considerations on GLCs. In [560], Yu and McMillan gave a projection equation for a special case of GLCs (concretely, a transformation that maps each 3D point to the ray it lies on in the GLC). They also showed how to relate two GLCs which differ in the choice of image plane. Whereas for perspective cameras, corresponding points in two different image planes (but with the same optical center) are related by a homography, this relationship is a quartic rational function for the considered special case of GLCs. In the general case of GLCs, this relation will be even more complicated. However, if parallel reference planes are used, then changing the distance between them does not matter: for any GLC with a particular distance between the planes, there is an equivalent GLC defined with planes separated by another distance.

The GLC model is defined by three basis rays. An extension to multiple basis rays was proposed and analyzed for up to six rays, by Popescu et al. [406].

3.4.5 Linear Families of Lines

Ponce proposed to model the set of camera rays by two-parameter linear families of straight lines, either degenerate reguli (rank-3 families) and non-degenerate linear congruences (rank-4 families) [405]. This framework subsumes the GLCs of Yu and McMillan [558, 559] and the linear oblique cameras of Pajdla [390]. Ponce gives formulas for projection, back-projection, and epipolar and multi-view geometry, formulated such that the two parameters of the considered linear families of camera rays are considered as image point coordinates. The back-projection matrix is of size 4×6 for two-slit cameras and linear oblique ones, and 5×6 for the "pencil cameras" of Yu and McMillan [559]. All models subsumed by Ponce's formulation can be expressed by 6×6 back-projection matrices, leading to 6×6 fundamental matrices (cf. Section 4). Batog et al. generalized the theoretical framework

of [405] by characterizing all admissible maps that allow to model linear cameras, i.e., cameras whose rays form a linear family of lines [43].

3.4.6 Caustics

Roughly speaking, the caustic of an imaging system is the surface enveloping the set of camera rays, i.e., it is tangent to all rays. Caustics are an excellent tool to analyze properties of imaging devices, as nicely advocated by Swaminathan et al. [484]. A caustic provides one possible description of a viewpoint locus of a camera, i.e., a point set such that any camera ray goes through at least one of the points in the set (ideally through exactly one). For a central camera, the caustic degenerates into a point and for axial ones, into a line segment. The caustic allows to study "how central" a camera is, i.e., how much its extent deviates from a single point. Swaminathan et al. showed how to compute the extent and shape of the caustic for axial catadioptric systems based on conic-shaped mirrors and such that the camera's optical center lies on the mirror axis [484]. They showed how to use this analysis to determine a best approximation of a single viewpoint, useful for example for approximate perspective view generation. They also analyzed how resolution varies across the image area, and reciprocally, proposed an approach to compute a mirror profile that satisfies a desired resolution profile. Finally, the parametric form of the caustic for a class of catadioptric sensors is used in a calibration approach, described in Section 5.3.2.

Usually, to each image point corresponds exactly one point on the caustic. This corresponds to the "viewpoint" associated with that image point. To compute the entire camera ray, one still needs its direction. For axial systems, it can be easily computed since the camera ray must pass through the axis and also be tangent to the caustic. For fully non-central systems, the full determination of camera rays requires a second map (of directions), besides a map from image points to the caustic.

3.5 Overview of Some Models

In Table 3.1, some of the models seen in this section are summarized. Note that in the table, we do not always use the same notation for these

Table 3.1. Back-projection and forward projection equations for some camera models.

Name of model	B^i	$\theta(r_d)$	$r_u(r_d)$	P^i	$r_d(\theta)$	$r_d(r_u)$
Perspective	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & f \end{pmatrix}$	$\operatorname{atan}\frac{r_d}{f}$	r_d	$\begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$f\tan\theta$	r_u
Classical radial dist.	$\begin{pmatrix} k & 0 & 1 & 0 & 0 \\ 0 & k & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & f \end{pmatrix} \begin{pmatrix} q_1^3 + q_1 q_2^2 \\ q_1^2 q_2 + q_2^3 \\ q_1 q_3^2 \\ q_2 q_3^2 \\ q_3^3 \end{pmatrix}$	$\operatorname{atan}\frac{r_d(1+kr_d^2)}{f}$	$r_d(1+kr_d^2)$	Requires solution of cubic equation		
Para-catadioptric	$\begin{pmatrix} 0 & 2k & 0 & 0 \\ 0 & 0 & 2k & 0 \\ -1 & 0 & 0 & k^2 \end{pmatrix} \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}$	$\operatorname{atan}\frac{2kr_d}{k^2 - r_d^2}$	N/A	Requires solution of quadratic equation		
Division [62, 145]	$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ kf & 0 & 0 & f \end{pmatrix} \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}$	$\operatorname{atan}\frac{r_d}{f(1+kr_d^2)}$	$\frac{r_d}{1+kr_d^2}$	N/A	N/A	N/A
Mičušik, Pajdla [344]	N/A	$\frac{r_d}{f(1+kr_d^2)}$	N/A	N/A	N/A	N/A
Stereographic	$\begin{pmatrix} 0 & 2f & 0 & 0 \\ 0 & 0 & 2f & 0 \\ -1 & 0 & 0 & f^2 \end{pmatrix} \begin{pmatrix} q_1^2 + q_2^2 \\ q_1 q_3 \\ q_2 q_3 \\ q_3^2 \end{pmatrix}$	$2\operatorname{atan}\frac{r_d}{f}$	N/A	N/A	$f\tan(\theta/2)$	N/A
Equidistant/-angular	N/A	$\frac{r_d}{f}$	N/A	N/A	$f\theta$	N/A
Sine-law	N/A	$\operatorname{asin}\frac{r_d}{f}$	N/A	N/A	$f\sin\theta$	N/A
Equi-solid angle	N/A	$2\operatorname{asin}\frac{r_d}{f}$	N/A	N/A	$f\sin(\theta/2)$	N/A
Fisheye transform [41]	N/A	$\operatorname{atan}\frac{e^{r_d/k}-1}{f}$	N/A	N/A	$k\log(1+f\tan\theta)$	N/A
Herbert [219]	N/A	N/A	N/A	N/A	$\theta + \sum_{i=1}^{5} a_i \sin\frac{i\pi\theta}{\theta_{\max}}$	N/A
Kumler and Bauer [294]	N/A	$\frac{\operatorname{asin}\frac{r_d}{f}}{k}$	N/A	N/A	$f\sin(k\theta)$	N/A
FOV model [113]	N/A	$\operatorname{atan}\frac{\tan(kr_d)}{2f\tan\frac{k}{2}}$	$\frac{\tan(kr_d)}{2\tan\frac{k}{2}}$	N/A	N/A	$\frac{1}{k}\operatorname{atan}(2r_u\tan\frac{k}{2})$
Bakstein and Pajdla [23]	N/A	N/A	N/A	N/A	$k_1\tan\frac{\theta}{k_2} + k_3\sin\frac{\theta}{k_4}$	N/A
Gennery [161]	N/A	N/A	N/A	N/A	$\frac{\sin(\beta\theta)}{k\cos(\max(0,\beta\theta))}$	N/A

models as in the previous sections; rather, we try to homogenize notations by naming coefficients of camera models k or $k_1, k_2, \ldots$. Further, parameters akin to the focal length are denoted by f.

The shown models are all radial distortion models; their equations are given under the assumption that the distortion center is identical to the principal point and located at the origin. How to relax these assumptions is outlined in the introduction to this section.

The table contains six different ways of modeling the imaging transformation. Cells annotated by "N/A" correspond to cases where there is no closed-form expression or, if there is no expression of back-projection in terms of polynomials (second column).

Sample graphs of the function $\theta(r_d)$ for several camera models are given in Figure 3.4.

3.6 So Many Models . . .

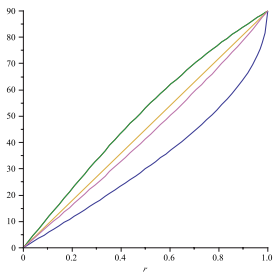
This section reviewed some camera models existing in the computer vision literature. Although it is certainly not exhaustive, the list of models is rather long. One may ask, which model one should choose or recommend. For example, recurring questions are, should one use a non-central camera model for a slightly misaligned catadioptric system or stick with a central model, or should one model fisheye lenses with a central or non-central model.

The general answer to these questions is: it depends . . .

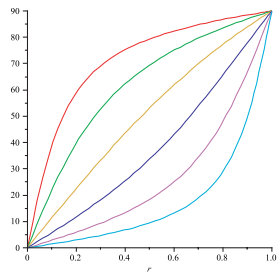
The choice of camera model depends of course on how appropriate a model is in general for the type of camera used. For example, one would obviously never use a pinhole model for fisheye cameras. More interesting parts of the question are: how to determine the appropriate order of a polynomial model, how to select among altogether different types of models, how to decide if it is worth using a non-central model rather than a central one, etc.

The choice of camera model or of the number of coefficients of a model may depend on several factors:

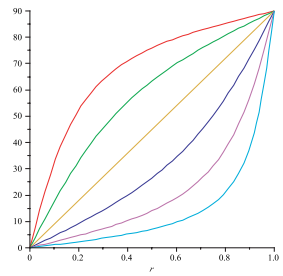
- The amount of noise in the calibration data.
- The amount of calibration data: how many images are used and how many calibration points (or other primitives used for calibration)?



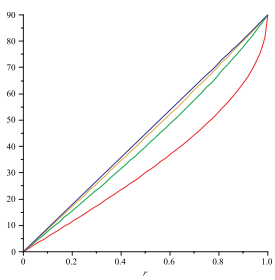
(a) Classical fisheye models. From left to right: stereographic (equal to para-catadioptric), equiangular, equi-solid angle, sine-law.



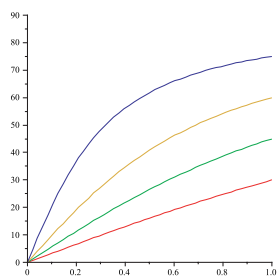
(b) Division model. From left to right, $f = \frac{1}{8}, \frac{1}{4}, \ldots, 4$.



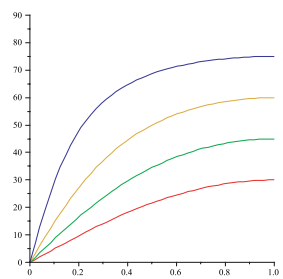
(c) FOV model. From left to right, $f = \frac{1}{8}, \frac{1}{4}, \ldots, 4$.



(d) Kumler-Bauer model. From left to right, $k = 4, 2, 1.25, 1$ (1 is the smallest possible value).



(e) Pinhole model. Focal lengths chosen such that at $r_d = 1$, we have $\theta = 30°, \ldots, 75°$.



(f) Classical radial distortion model with one coefficient. Coefficient k chosen such that θ is maximum for $r_d = 1$ ($k = -\frac{1}{3}$) and focal lengths chosen such that at $r_d = 1$, we have $\theta = 30°, \ldots, 75°$.

Fig. 3.4 Graphs of the function $\theta(r_d)$ for some camera models. In the first two rows, distortion model coefficients k are chosen such that for $r_d = 1$, we have $\theta = 90°$.

- The quality of the spatial distribution of calibration data: how well do the data cover the field of view (both across the field of view and in depth), how good is the spatial distribution of the set of camera positions, how different are the camera orientations, etc.?
- The "true" camera model: is it central or not, how pronounced are non-perspective distortions, etc.?

The choice depends on the combination of these and possibly other factors. For example, Zhang showed that the self-calibration of the classical radial distortion model, even with a single coefficient, from matches between images, is only stable if the actual distortion is pronounced **and** if extracted image points are accurate [566].

Let us consider extreme cases. Without noise in the data, the true model can be estimated without error (assuming we are not in a degenerate situation for the position of the calibration data, such as the twisted cubic for pinhole camera calibration [67]). Even with noise, the availability of infinitely many calibration data may allow a perfect calibration of the true camera model.

In practice, there is noise and the number of data is finite and we must consider the interplay of the above factors. To stress that the choice of camera model depends on all of these factors, we may note that even for one and the same camera, different models may be chosen for different scenarios. For example, fisheye lenses are probably not perfectly central, i.e., not all camera rays go through a same 3D point. However, the viewpoint locus is at most as large as the aperture of the lens, i.e., a few centimeters, and in practice, much smaller. Hence, if the fisheye lens is used for an application where the scene of interest is meters away from the camera, then an approximation by a central imaging model should be sufficient (unless we use many images and extremely accurate image points for calibration). However, if the scene in some hypothetical application is much less than a meter away, a non-central model may make sense.

So far, the discussion was rather qualitative. To actually implement a method for choosing among camera models requires quantitative goodness measures.

Clearly, our question is one of model selection, for which a large body of literature exists. The main principle of model selection is to find a good compromise between the goodness of fit to the data and the complexity of a model; roughly speaking, a good model should explain the data well but not have too many parameters. The reason for this is that in general, a model with more parameters has a better chance to

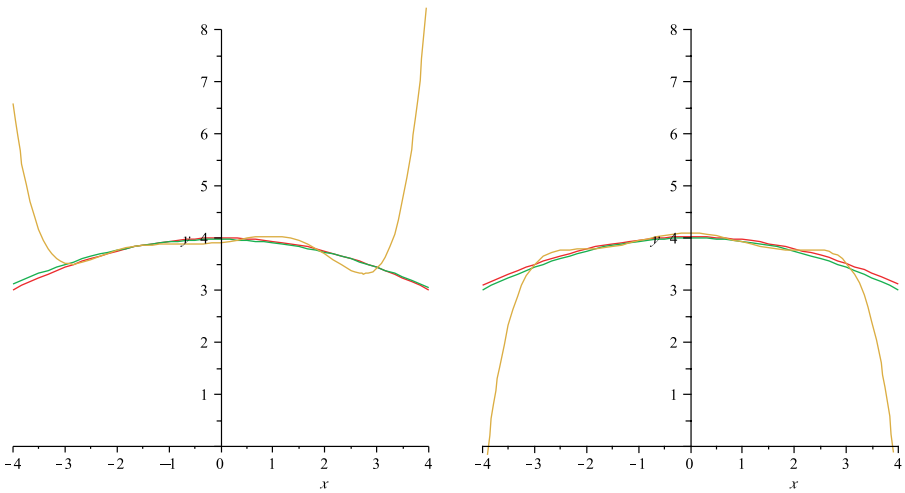


Fig. 3.5 Illustration of overfitting. Seven points are sampled from a quadratic polynomial (red curve) and perturbed by random noise. Then, a quadratic and a degree-6 polynomial are fitted to the points (green and orange curve respectively). The degree-6 polynomial fits the data perfectly as expected but is an extreme overfit as can be seen from the two random trials above, whereas the degree-2 polynomial fits stably and accurately.

well fit the data.[4] If more parameters than necessary are allocated then this constitutes an overparameterization and it may lead to overfitting: in addition to explaining the phenomenon underlying the observed data (here, an imaging process) the model parameters will tend to explain the noise in the data.[5] If we tip the balance among the above factors, e.g., too few or too badly distributed data compared to the amount of noise, then the model will no longer be a trustworthy explanation of the phenomenon. The usual sample illustration of overfitting concerns the problem of fitting a polynomial to data points, cf. Figure 3.5.

The problem of overfitting in camera calibration is real but can be rather safely reduced by following a few guidelines. If an application allows it, one should always acquire more than the minimum required

[4] The fit gets always better if one extends the number of parameters of the *same* model, e.g., by using additional coefficients for modeling radial distortion or a finer tessellation for local camera models. As for *different* models, this needs not always be the case.

[5] Another manifestation of this are correlations between the estimated parameters.

amount of data (in terms of number of images, number of calibration points, etc.). The image area should be completely covered by calibration points: distortions are usually strongest, i.e., most observable at the image border and even in the absence of distortions, a better coverage of the image area gives a better leverage on calibration and other structure-from-motion tasks (e.g., motion estimation gets stabler the wider the field of view that is effectively used). By design, calibration grids usually have well-distributed calibration points. One of the most important aspects is that images should be taken from as different as possible viewpoints and, crucially, with as different camera orientations as possible. The latter aspect includes rotations about the optical axis (required if an aspect ratio is to be computed) but more importantly, one should acquire images with different directions of the optical axis (or, orientations of the image plane). Naturally, one should also spend an effort to get the most accurate image point extraction.

These are just a few general guidelines. More intricate ones which may be important for high accuracy calibration are that cameras should best be calibrated in the same conditions as in the final application, e.g., same temperature, atmospheric conditions, even in the same orientation (tilting a camera may already affect its intrinsic parameters), etc. More on these issues can be found in photogrammetric literature, e.g., [336, 451].

Coming back to the model selection problem: model selection methods for camera calibration were proposed for example by El-Melegy and Farag [127], Ma et al. [327], and Orekhov et al. [387]. See also the good discussion by Derrien and Konolidge in [111], who also evaluated the geometric error committed when approximating a non-central camera by a central model. An alternative to usual model selection methods is to evaluate the quality of different camera models (and of the estimated coefficients thereof) using other data than that used for calibration. One possibility is cross-validation, i.e., evaluate calibration results using calibration images not used for calibration. Another one is to judge the quality of calibration within an actual intended application of the calibration: calibration is rarely a goal in its own but is carried out in order to realize some application, such as 3D modeling or egomotion estimation. If ground truth data for these or other applications

are available, then it is obviously a good idea to measure the quality of a calibration result via the quality of the subsequent result in the considered application.

A final comment is that while it is not possible to recommend this or that camera model in general, it seems a good idea to enhance "ideal" models for cameras with large fields of view (e.g., the equiangular one), with classical radial or tangential models, as proposed for example by Devernay and Faugeras [113], Kannala and Brandt [266], Gennery [161], and Mei and Rives [339].

4

Epipolar and Multi-view Geometry

Epipolar geometry is central to stereovision. On the one hand, knowing it enables to strongly constrain the problem of matching two images. On the other hand, it can be estimated from matches and then allows for motion estimation, self-calibration, and triangulation of 3D points or other geometric primitives. Epipolar geometry has been well studied for perspective cameras, since the nineteenth century (see, e.g., the excellent paper [144] by Finsterwalder on geometric foundations of photogrammetry, as well as modern treatments [137, 213]).

As for non-perspective cameras, much has been done in the last 20 years and this section aims at giving a concise overview. It will be seen in the following that epipolar geometry is a very powerful tool for the self-calibration of non-perspective distortions; this is based on a strong link to the plumb-line calibration approach (cf. Section 5.2.1), explained in Section 4.3.

We start the section by considering the case of calibrated cameras, where the epipolar geometry can be represented by an essential matrix. Except the question if cameras are central, non-central, axial, etc., the essential matrix is independent of the camera type. In Section 4.2, we consider the epipolar geometry of uncalibrated cameras

and particularly, fundamental matrices for non-perspective cameras. Finally, since epipolar curves are images of straight lines (camera rays of the other camera), we consider the computation of line images in Section 4.3 and, most importantly, the above mentioned strong links between self-calibration and plumb-line calibration of non-perspective cameras.

4.1 The Calibrated Case

The epipolar geometry for a pair of calibrated *central* cameras can be represented by the essential matrix E [324] and the associated epipolar constraint on back-projected image points:

$$(\mathbf{B}_2^i)^\mathsf{T} \underbrace{([\mathbf{t}]_\times \mathsf{R})}_{\mathsf{E}} \mathbf{B}_1^i = 0, \tag{4.1}$$

where $\mathbf{t}$ and R are the relative translation and rotation between the two views. Note that since we directly work with the directions of camera rays, the above formulation is independent of the type of central camera and also works when considering two cameras of different types.

The essential matrix for *non-central* cameras was introduced by Pless [402]. For non-central cameras, we must consider the complete camera rays instead of only their directions as in Equation (4.1). Let $\mathbf{L}_1$ and $\mathbf{L}_2$ be the camera rays associated with matching image points in a pair of views. They must intersect one another (in the 3D point whose images are the two matching image points). Putting together Equations (1.2) and (1.3) of Section 1.2, this leads to the essential matrix and epipolar constraint for non-central calibrated cameras:

$$\mathbf{L}_2^\mathsf{T} \begin{pmatrix} -[\mathbf{t}]_\times \mathsf{R} & \mathsf{R} \\ \mathsf{R} & 0 \end{pmatrix} \mathbf{L}_1 = 0. \tag{4.2}$$

The essential matrix acts on Plücker coordinates of camera rays and is thus of size 6×6.

In [402], Pless also gave the epipolar constraint for continuous camera motion. We refer to Section 6.2 for notes on the estimation of the central and non-central essential matrices and subsequent camera motion estimation.

Sturm specialized Pless' epipolar constraint to calibrated axial, two-slit, and central cameras, for all cases with finite/infinite slits respectively optical centers [472]. For axial and two-slit cameras, the essential matrices are of size 5×5 and 4×4, respectively. It was also shown how to establish trifocal matching tensors for these camera models.

Gasparini and Sturm showed how to derive matching tensors for line images as opposed to images of points, for central, two-slit, axial, and fully non-central calibrated cameras [158].

4.2 The Uncalibrated Case

4.2.1 Generalities

The epipolar geometry of uncalibrated perspective cameras can be represented by the fundamental matrix [137, 213]. How about other central cameras or the case of non-central ones? The answer is immediate if back-projection can be represented as a matrix–vector product (or, by polynomials or rational polynomials, cf. Section 3.1.8).

Central cameras. Let B^i be the back-projection matrix of a central camera, of size $3 \times n$ and operating on lifted image coordinates $\bar{\mathcal{L}}^a(\mathbf{q})$, where a and n depend on the camera model (e.g., $n = 5$ and $a = 3$ for the one-coefficient classical radial distortion model, cf. Section 3.1.1). By plugging $\mathbf{B}_1^i \sim \mathsf{B}^i \bar{\mathcal{L}}^a(\mathbf{q}_1)$ and $\mathbf{B}_2^i \sim \mathsf{B}^i \bar{\mathcal{L}}^a(\mathbf{q}_2)$ into the epipolar constraint (4.1), we directly get the fundamental matrix:

$$(\bar{\mathcal{L}}^a(\mathbf{q}_2))^\mathsf{T} \underbrace{(\mathsf{B}^i)^\mathsf{T} [\mathbf{t}]_\times \mathsf{R}\, \mathsf{B}^i}_{\mathsf{F}} \bar{\mathcal{L}}^a(\mathbf{q}_1) = 0.$$

It is of size $n \times n$.

It is of course possible to derive the fundamental matrix for two cameras of different types with back-projection matrices of different sizes $3 \times m$ respectively $3 \times n$:

$$(\bar{\mathcal{L}}^b(\mathbf{q}_2))^\mathsf{T} \underbrace{(\mathsf{B}_2^i)^\mathsf{T} [\mathbf{t}]_\times \mathsf{R}\, \mathsf{B}_1^i}_{\mathsf{F}} \bar{\mathcal{L}}^a(\mathbf{q}_1) = 0. \tag{4.3}$$

The fundamental matrix will accordingly be of size $m \times n$.

Due to the presence of the rank-2 skew-symmetric matrix $[\mathbf{t}]_\times$, the fundamental matrix is always of rank smaller or equal than 2, in general of rank 2.

Non-central cameras. The situation for non-central cameras is analogous: it suffices to plug back-projection matrices B^l of size $6 \times n$ into the non-central epipolar constraint (4.2). A difference from the central case is that the rank of the non-central fundamental matrix depends on the rank of the back-projection matrices as well as on the relative position of the two cameras. For example, the back-projection matrices of two-slit cameras are of rank 4; the rank of the two-slit fundamental matrix is 4 if the two views are in general position, and lower if the two views share one slit or if their slits are incident (cf. [139] and Section 4.2.2).

Epipolar curves. For perspective cameras, $\mathsf{F}\mathbf{q}_1$ represents the epipolar line in the second view, associated with a point $\mathbf{q}_1$ in the first view. In our general setting, epipolar lines are replaced by epipolar curves. It is easy to see that the order of epipolar curves is identical to the order of image coordinate lifting (a and b respectively in Equation (4.3)). For example, for liftings of order $2, 3$ respectively 4, epipolar curves are conics, cubic respectively quartic curves. This observation tells us immediately that for all cameras that can be modeled by back-projection matrices operating on second order lifted image coordinates, epipolar curves are conics, e.g.,: linear pushbroom cameras and other types of two slits (cf. Section 3.1.4), linear oblique cameras (cf. Section 3.4.3), para-catadioptric cameras (cf. Section 2.3.1), the division model (cf. Section 3.1.8), all GLCs (cf. Section 3.4.4), etc. Also, epipolar curves for the one-coefficient classical radial distortion model (cf. Section 3.1.1), are obviously cubic (cf. [566]).

As described in Section 3.4.2, non-central cameras may or may not be in standard stereo configuration: they are, if epipolar curves in the two images are in a one-to-one correspondence. This is for example the case for two-slit cameras with relative positions as mentioned above (e.g., they share a slit), which corresponds to a fundamental matrix of rank 2. Otherwise, one can still compute epipolar curves in both

images, but they now form 2D families in each image and are no longer in one-to-one correspondence. An algebraic consequence is an increase in the rank of the fundamental matrix, for the above example of two-slit cameras, an increase from 2 to 3 or 4.

Let us note that Rademacher and Bishop already described how to understand epipolar curves for arbitrary non-central images [411]. They also suggested the concept of an internal epipolar geometry, i.e., an epipolar geometry of a non-central camera with itself: when considering any ray of the camera, all other rays that cut it, are associated with an epipolar curve. For oblique cameras, no internal epipolar curves exist by definition.

Epipoles. For central cameras, one may define epipoles analogously to the pinhole case: the image of the optical center of one camera is the epipole in the other camera's image plane. Further, all epipolar curves go through the epipole. In general, there may be more than one epipole, for example for central catadioptric cameras (see an example in Figure 4.1) for which each 3D point is mathematically imaged in two points (cf. Section 3.1.5). In that case, all epipolar curves go through all epipoles. Another example is the classical radial distortion model;

(a) Perspective image. (b) Para-catadioptric image.

Fig. 4.1 Epipolar geometry for a perspective and para-catadioptric image pair. Cyan circles indicate three point matches and associated epipolar lines/curves are shown in red. Epipoles are shown in white; there exist two of them in the para-catadioptric image.

in the case of one distortion coefficient, there are three epipoles: either three real points or one real and two conjugate imaginary points.

For non-central cameras, the notion of optical center and thus of epipole does not exist anymore. However, in the special cases of axial and two-slit cameras, a weaker notion can be defined as follows: the image of the camera axis (or axes in the case of a two-slit camera) in the other camera gives a curve that is incident to all epipolar curves.

The non-polynomial case. The results in this section so far only apply to cameras whose back-projection can be modeled by matrices. In other cases, e.g., for most of the fisheye models of Section 3.1.7 (cf. also Table 3.1), fundamental matrices do not exist. The epipolar geometry is still defined of course, it just cannot be expressed in the usual matrix–vector style. The epipolar geometry has been studied for some of these cases, see the following section.

4.2.2 Existing Works

In the following, a few existing works on the epipolar geometry of non-perspective cameras are reviewed.

Classical polynomial distortion model. Zhang, in [566], studied the epipolar geometry of two radially distorted images and showed that epipolar curves are cubic if only one radial distortion coefficient is considered. He proposed a non-linear optimization method for estimating the distortion coefficients of the two images and the perspective fundamental matrix.

Central catadioptric cameras. Svoboda et al. studied the epipolar geometry for central catadioptric cameras [482, 479, 481]. They showed that epipolar curves are conics and gave formulas for computing these conics, from known camera calibration and motion. They also showed the existence of two epipoles per image, in general, which are incident with each epipolar conic.

Geyer and Daniilidis showed that the epipolar geometry of para-catadioptric cameras can be represented via a fundamental matrix [166]. It acts on lifted coordinates of image points and is of

size 4×4 if one assumes that images have square pixels (in the general case, it is of size 6×6 [471]). It can thus be estimated linearly from 15 or more point matches. Geyer and Stewénius proposed a minimal 9-point algorithm to compute the fundamental matrix between two para-catadioptric cameras, based on constructing a Gröbner basis of the polynomial equation system given by the point matches [172].

Geyer and Daniilidis [168] and independently, Sturm [471], showed how to systematically derive multi-view relationships for para-catadioptric cameras. Their approaches use a direct extension of the framework used for perspective images by Faugeras and Mourrain [138] and Triggs [511]. They allow to re-derive the above para-catadioptric fundamental matrix and in addition, trifocal or quadrifocal tensors. Sturm showed that using the same framework, one may obtain multi-view relations for any combination of para-catadioptric, perspective, or orthographic images [471]. It is clear from Section 4.2.1 that this easily generalizes to all types of camera models represented by back-projection matrices.

Other central catadioptric cameras than para-catadioptric ones, i.e., those based on hyperboloidal or ellipsoidal mirrors, are less nicely modeled. The main reason for this is that mathematically, each image point is back-projected to two camera rays (cf. Section 3.1.5). Although only one of them is physically plausible, it is not possible to dissociate the two rays while expressing them using polynomials in the image point coordinates. Sturm and Barreto showed that it is possible to represent the two rays by a single geometric object (a line complex) which can be represented using polynomials [473]. This enabled the definition of back-projection and fundamental matrices, the latter being of size 15×15.

Non-central catadioptric cameras. Cone-shaped mirrors usually lead to non-central catadioptric cameras (cf. Section 2.3.1). The epipolar geometry for such cameras, when restricted to planar motions, was derived by Yagi and Kawato [542]. As for two cameras mounted one on top of the other, with aligned mirror axes, the epipolar geometry is simple [456]: the mirror axes pierce the two image planes in the epipoles and the radial lines going through the epipoles, are epipolar lines. One

may see that this situation is always true for pairs of radially symmetric axial cameras with aligned camera axes.

Würz-Wessel proposed a numerical approach for computing epipolar curves in images of free-form mirror surfaces [538].

Central and non-central panoramas. Ishiguro et al. showed that in central panoramic images (see Section 2.1.1), 3D lines are imaged as sinusoidal curves [246, 247]. Consequently, epipolar curves are sinusoidal too, a result also derived by McMillan and Bishop [338] and Bunschoten and Kröse [68, 69]. Torii et al. studied the multi-view geometry of central panoramas [509].

Huang et al. studied the epipolar geometry for non-central panoramic images generated by a perspective 1D sensor rotated about an axis parallel to the sensor line but not necessarily passing through the 1D camera's optical center [236, 237, 238]. Further, the 1D sensor may be oriented such that the sensor line, the optical center, and the rotation axis are not coplanar. Huang et al. derived the epipolar geometry for such images as well as several special cases [236, 238] and proposed motion estimation methods [237]. Šivic and Pajdla proposed an in-depth geometric study of non-central panoramic images and their stereo configurations [450].

Menem and Pajdla established the epipolar geometry between a non-central panoramic image and a pinhole image [342].

Polynomial and rational polynomial models. Claus and Fitzgibbon uncovered the fundamental matrix for the bi-quadratic rational function model (see Section 3.1.8) [98]. It is a 6×6 matrix that maps lifted image points to the six coefficients (five plus a scale factor) of the epipolar conics in the other image. If one restricts the model to the one-parameter division model, then image lines are circles and epipolar geometry may be written via a 4×4 fundamental matrix [31, 98].

Barreto and Daniilidis showed that this is the case even if the two cameras have different distortion coefficients [31]. They also showed how to compute the fundamental matrix and extract the distortion coefficients from it. In [35], they showed that the combination of a para-catadioptric camera and one following the division model is also subject to a fundamental matrix.

Two-slit cameras. Gupta and Hartley showed that the epipolar geometry of linear pushbroom cameras can be modeled by a 4×4 fundamental matrix, expressed in lifted point coordinates [197]:

$$\begin{pmatrix} q_{2,1}q_{2,3} & q_{2,1}q_{2,2} & q_{2,2}q_{2,3} & q_{2,3}^2 \end{pmatrix} \mathsf{F} \begin{pmatrix} q_{1,1}q_{1,3} \\ q_{1,1}q_{1,2} \\ q_{1,2}q_{1,3} \\ q_{1,3}^2 \end{pmatrix} = 0,$$

where $q_{j,k}$ denotes the kth homogeneous coordinate of a point in image j. Images of straight lines are hyperbolas and consequently, so are epipolar curves.

Feldman et al. have developed the fundamental matrix for more general two-slit cameras [139]. To be precise, they considered the case where a two-slit image is constructed from images acquired by a perspective camera that translates by a constant speed parallel to the image plane, such that at each acquisition instant t, the vertical pixel column at coordinate $s(t) = \alpha t + \beta$ is taken from the perspective image, where α and β can be arbitrary values. The final image is obtained by concatenating these pixel columns. For any pair of values α and β, this construction corresponds to a two-slit camera. The first slit is the line of optical centers corresponding to the translational motion whereas the second slit depends on the values of α and β. The two slits, when seen along their common perpendicular, form a right angle.

Feldman et al. showed that two two-slit images constructed from the same original image sequence but with different values α and β are subject to an epipolar constraint embodied by a 6×6 fundamental matrix, acting on the lifted image coordinates. They also revealed the general conditions under which two two-slit images have an epipolar geometry (see also Section 3.4.2); this is the case if the two cameras share one slit completely or if each slit of one camera intersects both slits of the other camera. In the first case, the stereo system has a pencil of epipolar planes, in the second case, these are replaced by epipolar quadrics.

Note that even if the cameras are not in one of these configurations, there exists a fundamental matrix for two-slit images, of size 6×6. It allows, as usual, to compute the epipolar curve (here, a conic) for every

image point. However, epipolar curves in the two images are no more in a one-to-one correspondence: different points on the same epipolar curve in one image, may be associated with different epipolar curves in the other image.

Khan et al. showed that there exists a fundamental matrix between a linear pushbroom panorama and a perspective image and used this for estimating the position of a query perspective image with respect to a set of georeferenced panoramic images [273].

One-dimensional radial camera model. Thirthala and Pollefeys showed the existence of quadrifocal tensors for the 1D radial camera model [500]; refer to Section 3.1.9 for some more details.

The case of pure rotations. It is well known that two images taken by a pinhole camera undergoing a pure rotation about its optical center are related by a homography. How about other central cameras?

We consider the case of cameras modeled by back-projection matrices of size $3 \times n$, acting on a-order lifted image coordinates. For any pair of matching points $\mathbf{q_1}$ and $\mathbf{q_2}$, we must have:

$$\mathsf{R}\,\mathsf{B}_1^i\,\bar{\mathcal{L}}^a(\mathbf{q_1}) \sim \mathsf{B}_2^i\,\bar{\mathcal{L}}^a(\mathbf{q_2}),$$

where R is the rotation relating the two images. This constraint can be written as:

$$\{\mathsf{R}\,\mathsf{B}_1^i\bar{\mathcal{L}}^a(\mathbf{q_1})\} \times \{\mathsf{B}_2^i\bar{\mathcal{L}}^a(\mathbf{q_2})\} = \mathbf{0}_3,$$

which is equivalent to three bilinear constraints (only two being independent) on the lifted image coordinate vectors:

$$(\bar{\mathcal{L}}^a(\mathbf{q_2}))^\mathsf{T}\mathsf{M}_j\bar{\mathcal{L}}^a(\mathbf{q_1}) = 0, \quad j = 1\cdots 3,$$

with three matrices M_j of size $n \times n$ which depend on the rotation matrix R and the two back-projection matrices.

There is no bijective homography in general, unlike in the pinhole case, but it is replaced by the above bilinear constraints, which allow, for any point $\mathbf{q_1}$, to compute all potential matches $\mathbf{q_2}$. In general, there are only finitely many matches. To be precise, since the matching constraints are linear in the lifted coordinates of $\mathbf{q_2}$, there are in general

at most a^2 potential matches (two independent constraints on a-order lifted coordinates).

As for the 1D radial camera model and the case of pure rotations, Thirthala and Pollefeys showed the existence of trifocal tensors [500, 501] (cf. Section 3.1.9).

The case of planar scenes. Like for pinhole cameras, the case of a planar scene is very similar to that of a rotating central camera. For example, Draréni et al. derived the expression for plane homographies for linear pushbroom cameras and proposed a calibration approach using planar grids [120]. Fitzgibbon mentioned plane homographies for the division model [145]. Sturm and Barreto showed the existence of plane homographies for all central catadioptric cameras [473].

4.3 Images of Lines and the Link between Plumb-line Calibration and Self-calibration of Non-perspective Cameras

As will be discussed in Section 5.2.1, non-perspective camera parameters can in general be fully calibrated from images of straight lines. There is a strong link between this, epipolar geometry and self-calibration: from point matches between images, one can compute the epipolar geometry. Epipolar geometry allows to compute epipolar curves, which are nothing else than images of straight lines (of camera rays of the respective other camera). It is thus obvious that epipolar geometry allows for the self-calibration of non-perspective imaging models, at least partially. One may even directly use any plumb-line technique for self-calibration: from the epipolar geometry, one may generate arbitrarily many epipolar curves and use these as input to plumb-line calibration. This direct link between plumb-line calibration, epipolar geometry, and self-calibration is implicitly present in all non-perspective self-calibration approaches, although it has not been explained before in its full generality. It will be developed further in Section 5.3.1.

Line images. As for the computation of a 3D line's image, this is particularly simple for cameras modeled by back-projection matrices.

In the case of non-central cameras and back-projection matrices of size $6 \times n$, operating on a-order lifted image coordinates $\mathcal{L}^a(\mathbf{q})$, the image of a 3D line $\mathbf{L}$ is computed as follows. An image point $\mathbf{q}$ lies on the line image if its back-projected camera ray intersects the line $\mathbf{L}$, i.e., if (cf. Equation (1.3)):

$$\underbrace{\mathbf{L}^\mathsf{T} \begin{pmatrix} \mathsf{0}_{3\times3} & \mathsf{I}_{3\times3} \\ \mathsf{I}_{3\times3} & \mathsf{0}_{3\times3} \end{pmatrix} \mathsf{B}^l}_{\mathbf{c}^\mathsf{T}} \mathcal{L}^a(\mathbf{q}) = 0. \tag{4.4}$$

We immediately see that the image of $\mathbf{L}$ is the curve of order a whose coefficients are given in the vector $\mathbf{c}$. In the case of $a = 2$ for example, the line image is a conic; the six coefficients defining its symmetric matrix representation, are contained in $\mathbf{c}$. An example of a line image in a linear pushbroom panorama is shown in Figure 4.2.

In the case of central cameras, we can obtain the analogous result after computing the full back-projection ray from $\mathbf{B}^f$ (the origin) and $\mathbf{B}^i \sim \mathsf{B}^i \mathcal{L}^a(\mathbf{q})$ (the direction of the ray):

$$\begin{pmatrix} \mathsf{B}^i \mathcal{L}^a(\mathbf{q}) \\ \mathbf{0} \end{pmatrix} \sim \begin{pmatrix} \mathsf{B}^i \\ \mathsf{0} \end{pmatrix}_{6\times n} \mathcal{L}^a(\mathbf{q}).$$

By inserting this in Equation (4.4), we compute the line image for central cameras as:

$$\mathbf{c}^\mathsf{T} \sim \mathbf{L}^\mathsf{T} \begin{pmatrix} \mathsf{0}_{3\times3} & \mathsf{I}_{3\times3} \\ \mathsf{I}_{3\times3} & \mathsf{0}_{3\times3} \end{pmatrix} \begin{pmatrix} \mathsf{B}^i \\ \mathsf{0} \end{pmatrix} \sim \mathbf{L}^\mathsf{T} \begin{pmatrix} \mathsf{0} \\ \mathsf{B}^i \end{pmatrix}.$$

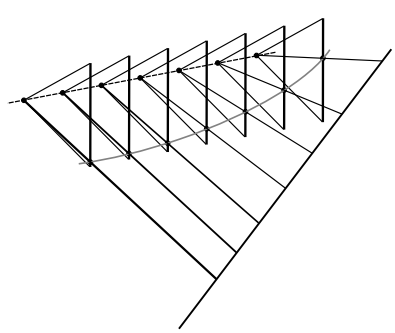


Fig. 4.2 The image of a line in a linear pushbroom panorama.

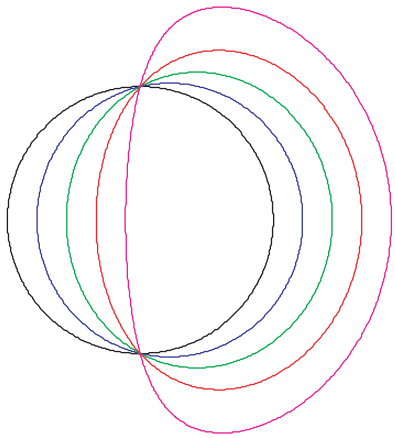


Fig. 4.3 Line images for the equiangular camera model. The black curve is the image of any line in the focal plane i.e., the plane containing the optical center and perpendicular to the optical axis. It is naturally a circle and corresponds to the hemispheric part of the field of view. The other curves show the images of lines whose interpretation planes (plane spanned by a line and the optical center) form angles of $70, 50, 30$, and 10 degrees with the optical axis, respectively. The limiting case of 0 degrees corresponds to lines that are coplanar with the optical axis and whose images are lines going through the distortion center. Although in general, line images are not algebraic curves, the parts within the hemispheric field of view, i.e., within the black circle, can be relatively well approximated by conics.

Let us denote by $\mathbf{l}$ the second sub-vector of length 3 of $\mathbf{L}$. This represents the line at infinity of the plane spanned by the optical center (the origin) and $\mathbf{L}$. The line image can then also be written as:

$$\mathbf{c}^\mathsf{T} \sim \mathbf{l}^\mathsf{T} \mathsf{B}^i.$$

Non-polynomial back-projection. In this case, line images cannot be computed by matrix–vector products. Still, it suffices to plug into Equation (4.4) the expression of back-projected image points to get an equation defining the line image, allowing for example to plot it. For example, Figure 4.3 shows line images for the equiangular camera model; the figure suggests that although they are not algebraic curves, e.g., are not conics, they can be well approximated by conics within a hemispheric field of view.

5

Calibration Approaches

This section aims at providing an overview of different calibration approaches, mainly with respect to the type of images or other information that are used for calibration, with an emphasis on approaches developed for omnidirectional cameras. The section is a far from exhaustive treatment of the topic of calibration. A few other references are as follows. An excellent overview of photogrammetric calibration approaches and their historical development, together with the development of distortion models, is the paper [95] by Clarke and Fryer, see also a paper by Remondino and Fraser [426] and a book section by Boland [55]. Among the many books covering the topic, we may cite the Manual of Photogrammetry, its fourth edition of 1980 [451] and the more recent fifth edition [336]. The article [430] by Roelofs is an interesting extended account of the state of the art on calibration up to 1951. Other relevant books are [19, 195, 213, 287] to cite just a few.

5.1 Calibration Using Calibration Grids

5.1.1 Three-Dimensional Grids

This is probably the most well-known calibration approach and will not be detailed much further. Let us just note that in order to calibrate omnidirectional cameras, several researchers used hollow

3D calibration objects painted with markers or patterns on the inside, such that cameras are put inside the object for calibration, see, e.g., [23, 61, 83, 413, 455, 506, 519, 552]. Usually, the object was used to gather 3D-to-2D matches which could then be used like in traditional calibration approaches. Ying and Hu used two hemispherical calibration objects with different radius to calibrate a model akin to the two-plane one (cf. Section 3.1.3) [552].

Among the many other approaches, let us mention one by Shih et al. who proposed a method for calibrating one radial distortion coefficient as well as the perspective camera parameters, via the solution of an 8×8 eigenvalue problem [448]. This is very similar to recent approaches for the simultaneous estimation of multi-view geometry (e.g., the fundamental matrix) and a distortion model (e.g., the division model), for example the approaches of Fitzgibbon and Claus [98, 145] and Mičušik and Pajdla [344, 346, 347, 349]. Shih et al. also gave a similar calibration method for the usage of planar grids.

Bastanlar et al. developed a calibration approach for central catadioptric cameras that is akin to the DLT approach for pinhole cameras [40]: computation of a projection matrix from 3D-to-2D matches, followed by the extraction of intrinsic parameters. In the case of catadioptric cameras, the projection matrix is somewhat different from that of a pinhole camera: as shown by Sturm and Barreto [473], it is possible to formulate a projection matrix but which does not directly map 3D points to 2D points, the reason being the existence of two theoretical image points (cf. Section 3.1.5). Instead, the projection matrix maps 3D points to a degenerate dual conic, representing these two image points. Besides this difference, DLT-like calibration is also possible, as exploited by Bastanlar et al. [40]. Gasparini et al. and Barreto et al. developed similar procedures for using planar grids [36, 159].

Huang et al. proposed a calibration approach for non-central panoramic images and compared it experimentally to a line-based and a self-calibration method [239].

5.1.2 Planar Grids

Many approaches have been developed for using planar grids in camera calibration, due to their ease of manufacture, storage, and use.

Planar or near-planar objects, e.g., crosses, are routinely used in photogrammetry, usually via a (self-calibrating) bundle adjustment procedure. Computer vision researchers proposed "closed-form" type solutions [475, 567], which are suboptimal and should be followed by a bundle adjustment (cf. Section 6.4).

In this section, we concentrate on plane-based calibration approaches developed for omnidirectional cameras or other cameras with general distortion profiles. There are many approaches for global camera models, which usually apply similar procedures: an optional initialization procedure, followed by a bundle adjustment. We will thus not describe them in any more detail and refer the reader to the references on camera models in Section 3.1. As for the local models, the calibration procedures designed for them are mostly already outlined in Section 3.2. In the following, we give a few details on plane-based approaches for discrete camera models, for two variants thereof — ray-based models and discrete samples of the distortion curve.

Ray-based calibration. Several similar approaches for ray-based or raxel-based calibration were proposed, e.g., by Gremban et al. [189], Champleboux et al. [85], and Grossberg and Nayar [191]. The principle is simple: two or more images of planar calibration grids are acquired, with a known motion of the camera or grid between the acquisitions. A dense matching between the grid in each position, and the camera image, is carried out. Hence, for each pixel, one can compute its camera ray by fitting a 3D line to the set of matched grid points, which can be put in the same coordinate frame since the motion between image acquisitions is known. This way, camera rays can be computed for each individual pixel (or some other spatial sampling of the image area, be it sub-pixel or super-pixel-wise).

Debaecker et al. used a similar procedure to calibrate their cone-based camera model from multiple images of planar grids [108]. These methods can be considered as dense versions of the two-plane approach, cf. Section 3.1.3.

This type of approach was generalized by Sturm and Ramalingam in order to allow an image acquisition without having to know the camera/grid motion [477]. While the approach based on the knowledge of camera/grid motion can in theory work with two images, a minimum

of three images is required in the case of unknown motion. Sturm and Ramalingam showed that from matches for three images, the unknown camera/grid motion can be recovered via the computation of particular trifocal tensors. Once the motion is known, the above approach can be readily applied to compute the camera rays and thus to finalize the calibration. In [477], the minimal case of three images was considered; this was extended toward using multiple images in [418], allowing for an easier calibration of omnidirectional cameras. Sturm and Ramalingam developed several variants of their approach, for using planar as well as 3D calibration grids, as well as for calibrating fully non-central [477], central [477], or axial cameras [420]. In [416], they gave a minimal solution for the calibration of a central camera using matches for four pixels and three grids.

Gonçalves and Araújo used this approach to first compute pixel–ray matches, to which they then fitted a catadioptric model [180].

Dunne et al. used a structured-light type approach where an LCD screen is used as planar calibration grid, allowing to achieve dense matches between the grid and the image through the display of a sequence of particular patterns [125]. Matches between the grid and one of the input images are used to define a warping from the image to the grid. The warping, when applied to that input image, will by definition lead to a perspectively correct image of the grid — the grid itself. This warping thus eliminates all non-perspective distortions. After applying it to other images of grids (they do not need to be of the same grid), one may use any calibration approach for perspective cameras to calibrate the perspective part of the camera projection, e.g., the methods of Zhang or Sturm and Maybank for planar grids [475, 567]. This approach combines the ideas of warp-based approaches (cf. Section 3.2.2) and ray-based calibration.

Discrete sampling of the distortion curve. In the 1950s, Hallert developed formulas for non-linearly optimizing either intrinsic or extrinsic parameters of a camera looking at a planar grid, supposing the respective other set of parameters is known [201]. He also included the computation of a radial distortion model in his procedure. The model he used is of the discrete sampling type (cf. Section 3.3.2): he

used a calibration grid of 5×5 regularly distributed points and seemed to assume that the grid can be positioned such that the center point coincides with the distortion center. Then, he computed the radial distance in the undistorted image, for every circle centered in the grid's center point and going through other points in the grid. From the computed values, he seemed to interpolate the complete distortion function, between distorted and undistorted radial distances. This is probably one of the first examples where radial distortion is determined by computing samples of the distortion function, as also done for example by Hartley and Kang [209] and Tardif et al. [491], see below.

Hartley and Kang proposed a set of methods for calibrating cameras with general radial distortion [209]. Their key observation is that point matches between a planar calibration grid and an image with whatever radial distortion are subject to a constraint akin to the epipolar constraint and that the epipole in the image plane is nothing else than the center of radial distortion. The latter may thus be computed by fitting a fundamental matrix to the point matches and extracting the epipole from it (one may also use multiple images and fundamental matrices to get a better estimate). It was shown that this can also be done, in a similar fashion, when using a 3D calibration grid or by self-calibration, i.e., from images of an unknown scene (see Section 5.3.1). Hartley and Kang then used the assumptions of a radially symmetric camera and of a monotonic distortion curve, to compute the latter, from multiple images of grids. They showed how to compute a discrete sampling of the distortion curve as well as how to fit parametric models thereof.

Hartley and Kang's approach was studied experimentally by Barreto et al. for the task of calibrating endoscopic cameras using planar grids [37]. It was found to be superior to using the division or a classical model for radial distortion.

Tardif et al. proposed two approaches for plane-based calibration of a general radially symmetric radial distortion model (cf. Section 3.3.2) [491]. The approach is similar in spirit to Hartley and Kang's although the method for finding the distortion center is less elegant. The basic versions of these approaches work for central cameras; an extension for non-central ones was also proposed.

5.1.3 One-Dimensional Grids: Sticks or Circular Grids Positioned in Radial Planes

Adorni et al. proposed a calibration method for the task of back-projecting an omnidirectional image onto a fronto-parallel ground plane [3]. They made the assumption of a radially symmetric camera and acquired an image of a linear calibration object lying on the ground plane, carefully positioned such that the camera's optical axis "pierces" the object at a known position. For each marker on the calibration object, the distance to the optical axis as well as the distance of the image point from the assumed distortion center is recorded in a look-up table. By interpolating the information contained in this look-up table and using the assumption of radial symmetry, each image point can be mapped to the point on the ground plane that it sees, and vice versa, i.e., back-projecting the image to the ground plane is possible. A similar approach was proposed earlier by Southwell et al. which instead of markers on the calibration object used one with a fractal Grey code pattern printed on it [455].

As for classical cameras, i.e., non-omnidirectional ones, using sticks with markers is probably the *de facto* standard approach for calibrating commercial multi-camera systems for motion capture, although many alternative approaches exist, e.g., [33, 480]. How to calibrate a single camera using a 1D object, was shown by Zhang [568].

5.1.4 Calibration Grids at Infinity — Collimators, Stellar Calibration, Vanishing Points

A classical photogrammetric calibration technique exploits arrays of collimators or devices with a movable collimator [95, 202, 269, 492]. The incidence angle of each collimator with the lens to be calibrated is measured, e.g., using theodolites or a goniometric apparatus. One can interpret such an array as a calibration grid lying at infinity: instead of providing matches between image positions and 3D points, they directly provide directions of camera rays for image positions. If a planar array is used (collimator beams are coplanar) and positioned such that that plane contains the distortion center, one can directly sample the radial distortion function from the observations, or fit any radial

distortion model to them. In the case of a non-planar array, tangential distortions may be measured as well. More generally speaking, one may fit any global or local central camera model to data obtained from collimators.

Another classical method termed stellar calibration consisted in exploiting the well-known angular position of stars to calibrate cameras (care had to be taken though to take into account, e.g., atmospheric refractions) [151]. This may again be seen as using a calibration grid lying at infinity.

Finally, we may mention calibration techniques exploiting vanishing points arising from scene structure, e.g., associated with parallel edges of buildings. Knowing the angles between the associated directions in 3D gives essentially the same input as collimators and images of stars, although usually with (much) lower accuracy. Also, the number of vanishing points in typical scenes is low, often not higher than three. Thus, this kind of observation could typically not be used to measure non-perspective distortions. However, from three vanishing points with known angles, up to three intrinsic parameters of the pinhole model can be estimated. Methods doing so are for example those by Gracie [185], Caprile and Torre [81], and Echigo [126].

Hughes et al. showed that vanishing points arising from sets of parallel lines may be used to calibrate the distortion center of radially symmetric distortion models [241]. They did this for the division model (which was used as an approximation to the equiangular model), by showing the following properties. Lines are imaged to circles and a set of parallel lines are imaged to circles which intersect in two vanishing points; these correspond to the two mathematical images under the division model, of the point at infinity associated with the set of lines. The line joining the two vanishing points also contains the center of radial distortion. Thus, when observing at least two different sets of parallel lines, the distortion center can be computed in a straightforward manner. It seems that this approach may be used for any radially symmetric distortion model. In [240], Hughes et al. extended the previous approach by estimating another intrinsic parameter besides the distortion center, namely the proportionality factor k of the equiangular camera model (cf. Section 3.1.7).

5.2 Using Images of Individual Geometric Primitives

By *individual* primitives, we mean that their relative position is not known.

5.2.1 Lines — The Plumb-Line Approach

Brown introduced the so-called plumb-line approach which allows to calibrate non-perspective distortions from images of straight lines [64]. Since then, many such approaches have been proposed for different camera models, see below. Line-based calibration approaches are attractive since they allow to separate perspective from non-perspective parts of the imaging model and since their algebraic complexity is typically lower than that of full calibration formulations, allowing for easier closed-form type solutions. Further, since lines are omnipresent in urban or interior scenes or in man-made objects, they allow for a calibration in many applications where calibration grids cannot be employed.

Classical polynomial distortion model. Kang proposed a plumb-line method where snakes that are constrained by the radial distortion model are used to fit to images of straight lines, thereby estimating the distortion coefficients [263].

Alvarez et al. [14] used a classical distortion model and a plumb-line type method, based on observations of sets of collinear points. They defined a cost function which measures how close undistorted points are from being collinear. That measure is not based on the Euclidean distance between points and lines, i.e., is an algebraic distance. However, Alvarez et al. showed that it is possible to find the distortion coefficients that are globally optimal under that measure, using a standard resultant-based technique. This is possible for up to two distortion coefficients but hints on how to possibly handle more than two are also given in [14].

Many other plumb-line approaches were proposed for the classical as well as other distortion models, e.g., [127, 240, 241, 272, 337, 408, 427, 485, 503, 510, 520, 521, 553, 575]. Ricolfe-Viala and Sánchez-Salmerón combined a plumb-line constraint with a constraint expressing the

invariance of the cross-ratio of collinear points, to calibrate a classical distortion model from images of planar grids [428].

Central panoramas. Smadja et al. proposed a plumb-line calibration method for circular central panoramas, acquired by a 1D camera rotating about its optical center [452].

Catadioptric cameras. As for catadioptric cameras, Geyer and Daniilidis introduced a plumb-line method [164, 165, 167]. Based on their unified model (see Section 3.1.5), they showed that from two respectively three images of lines, one may fully calibrate a central hyper-catadioptric respectively para-catadioptric camera. They gave an algorithm for the para-catadioptric case; Barreto and Araújo developed an algorithm for hyper-catadioptric cameras [28, 31] as well as an improved algorithm for para-catadioptric ones that is able to fully compute the mirror shape [30, 32]. Ying and Hu, Vandeportaele et al., and Wu et al. gave similar algorithms [518, 537, 551]. Previously, Geyer and Daniilidis also proposed a method for calibrating para-catadioptric cameras from images of sets of parallel lines [163].

It is noteworthy that central catadioptric cameras can be fully calibrated from just images of lines, contrary to perspective cameras, with or without the classical radial or tangential distortion models. In the latter case, images of straight lines only allow to calibrate the camera up to an arbitrary perspective projection, whereas for catadioptric cameras, a full calibration is possible. One explanation for this is that catadioptric cameras cannot be modeled by a perspective projection followed by a distortion in the image plane. Rather, the distortion already happens in the first step of the projection (in the unified model of Geyer and Daniilidis, when mapping 3D points onto a sphere). Every camera model for which this can be said may be fully calibrated from images of lines even if their relative position is unknown.

Fish-eye models. Devernay and Faugeras presented a plumb-line method in [113] and applied it to different distortion models: the classical backward model used by Brown [64], but also the models proposed by Basu and Licardie and the field-of-view model (see Section 3.1.7).

Polynomial and rational polynomial models. Claus and Fitzgibbon proposed two plumb-line methods for the bi-quadratic rational function camera model (see Section 3.1.8) [97]. They are both based on the fact that with that model, lines are imaged as conics. The first method starts with estimating conics from extracted line images and recovers distortion parameters via the factorization of a measurement matrix consisting of the matrices of all image conics. The second method is of the bundle adjustment type; it estimates the distortion model's coefficients using as cost function the Sampson distance between image points and line images (conics). Barreto as well as Strand and Hayman proposed similar plumb-line methods for the 1-parameter division model, for which line images are circles [26, 466].

Discrete camera models. Tardif et al. showed how to calibrate a general radially symmetric radial distortion model with a plumb-line approach [489]. Their method is applicable to both, a parametric radial distortion curve and a discrete sampling of it.

General remarks. Whenever the back-projection operation can be written using polynomials, it is easy to set up "closed-form" solutions for plumb-line calibration. Consider the case of back-projection matrices for central cameras, of size $3 \times n$, for any n. As we have seen in Section 4.3, the image of a line $\mathbf{L}$ can be written as:

$$\mathbf{c}^\mathsf{T} \sim \mathbf{l}^\mathsf{T} \mathsf{B}^i,$$

where $\mathbf{l}$ is the line at infinity of the plane spanned by the optical center and the line $\mathbf{L}$. If we have m line images, we can set up a system like:

$$\begin{pmatrix} \mathbf{c}_1^\mathsf{T} \\ \vdots \\ \mathbf{c}_m^\mathsf{T} \end{pmatrix} \sim \begin{pmatrix} \mathbf{l}_1^\mathsf{T} \\ \vdots \\ \mathbf{l}_m^\mathsf{T} \end{pmatrix} \mathsf{B}^i.$$

It is easy to see that the back-projection matrix can be recovered by factorizating the matrix on the left, like in factorization-based methods for structure-from-motion [508]. It can only be determined up to the pre-multiplication with an arbitrary 3×3 matrix, which

corresponds to a projective transformation of the directions of back-projected points — the perspective part of the imaging process. In other words, the non-perspective distortions are fully recovered. Once this is done, one may apply any calibration or self-calibration method designed for perspective cameras, to make the calibration complete.

This approach can work with a minimum of three line images, although this is not recommended in practice; in that case, the matrix on the left of the above equation can directly be adopted as solution for B^i. If prior knowledge on the camera is available, i.e., if constraints on B^i are available, fewer line images may be sufficient in theory.

One may apply the same approach to non-central cameras, with more general back-projection matrices, e.g., of size $6 \times n$. In that case however, it must be taken into account that often, they will be of rank lower than 6, a fact that has to be taken into consideration when estimating them.

Finally, as always, the best approach for calibration is to optimize a meaningful geometrical/statistical error criterion, possibly after initializing with any of the above approaches, which usually comes down to a non-linear bundle adjustment like optimization process. For example, one may simultaneously estimate the coefficients of undistorted lines and of camera model parameters where the cost function is the sum of Euclidean distances between undistorted lines and points. It will in general be better even to minimize distances in the original, distorted image, i.e., distances between original image points and line images, where the latter are parameterized by camera model parameters and parameters for line images (two for central cameras, four for non-central ones).

5.2.2 Spheres

Algorithms using images of spheres for the calibration of perspective cameras were proposed for example by Penna [398], Daucher et al. [104], Teramoto and Xu [498], and Agrawal and Davis [6], to name a few.

Stevenson and Fleck used images of spheres to calibrate a local camera model, see Section 3.2.3 [463].

Ying and Hu showed how to use sphere images for calibrating central catadioptric cameras [551]. In these cameras, spheres are mapped to

conics, like in perspective images (to be precise, the mathematical image of a sphere is the union of two conics, but in true cameras only one of them is usually visible). Ying and Hu showed that from three or more extracted conics, the catadioptric camera can be calibrated. Ying and Zha analyzed the relationships between sphere images and the image of the absolute conic and proposed a linear calibration algorithm based on these [554, 555, 556].

5.2.3 Circles

Herbert proposed a calibration procedure where instead of point matches or images of lines, projections of circles painted on the inside of a hemisphere are used [219]. Coefficients of his distortion models (see Section 3.1.7) are then estimated by measuring and comparing the imaged areas of such circles.

Many algorithms using circles for calibrating pinhole cameras have been proposed in the literature, see [91, 199, 257, 476] to name just a few.

5.3 Self-calibration

Self-calibration has attracted great interest in the computer vision community in the last close to 20 years. The possibility of camera self-calibration was already well known in photogrammetry, although it was usually identified with the non-linear simultaneous optimization of intrinsic and extrinsic camera parameters and scene structure (self-calibrating bundle adjustment), whereas computer vision researchers aimed at developing initialization approaches.

Most computer vision self-calibration works concern pinhole cameras; a good general reference is the book by Hartley and Zisserman [213]. In the following, we describe approaches for non-perspective cameras.

5.3.1 Self-calibration from Image Matches

General observations. As explained in Section 4.3, there is a strong link between plumb-line calibration (calibration from line images) and

self-calibration. From image matches, epipolar geometry can be recovered; epipolar curves can be generated and since they are images of straight lines (camera rays) they can be used as input to any plumb-line approach. Instead of this, mainly conceptual, idea of explicitly generating line images from a computed epipolar geometry, one may directly use fundamental matrices for self-calibration.

In the following, we only consider the self-calibration of the non-perspective part of a central camera model. By this we mean that it suffices to recover its $3 \times n$ back-projection matrix up to the pre-multiplication by any (invertible) 3×3 matrix M. The effect of M is a projective transformation of the directions of back-projected camera rays, i.e., a perspective part of the camera model.

In the following, we consider the case of a fundamental matrix between two images taken by the same central camera. Let B be its $3 \times n$ back-projection matrix (we omit the superscript i here for ease of notation). The fundamental matrix is then given by (cf. Equation (4.3)):

$$\mathsf{F} \sim \mathsf{B}^\mathsf{T}[\mathbf{t}]_\times \mathsf{R}\, \mathsf{B}.$$

Consider rows of F; they are linear combinations of the rows of B, with coefficients of these combinations given by the matrix $(\mathsf{B}^\mathsf{T}[\mathbf{t}]_\times \mathsf{R})$. Similarly, the columns of F are linear combinations of the columns of B^T, i.e., of the rows of B. This suggests the idea of determining the rows of B by computing a basis of the space of row and column vectors of F. To do so, let us now look at the singular value decomposition (SVD) of F:

$$\mathsf{F} \sim \mathsf{U}\Sigma\mathsf{V}^\mathsf{T},$$

where all matrices are of size $n \times n$. Since F is of rank 2 in general, the SVD can be reduced by only considering the first two columns of U, the first two rows of V^T and the two non-zero singular values in Σ:

$$\mathsf{F} \sim \bar{\mathsf{U}}_{n\times 2}\bar{\Sigma}_{2\times 2}\bar{\mathsf{V}}^\mathsf{T}_{2\times n}.$$

The columns of $\bar{\mathsf{U}}$ as well as the rows of $\bar{\mathsf{V}}^\mathsf{T}$ are linearly dependent on the rows of B. Let us stack them together in a $4 \times n$ matrix and

compute its SVD:

$$\begin{pmatrix} \bar{\mathsf{U}}^{\mathsf{T}} \\ \bar{\mathsf{V}}^{\mathsf{T}} \end{pmatrix}_{4\times n} = \mathsf{W}_{4\times n}\Gamma_{n\times n}\mathsf{X}^{\mathsf{T}}_{n\times n}.$$

In the absence of noise, the above matrix is of rank 3 in general, since all its rows are linearly dependent on the three rows of B. The three rows of X^{T} corresponding to the non-zero singular values in Γ can now be directly adopted as the rows of the back-projection matrix B. It is clear that the rows of B are only defined up to linear combination of the rows of X^{T}, but this does not matter: as stated above, B needs only be recovered up to a projective transformation M, and this means nothing else than a linear combination of its rows.

A more compact alternative procedure would proceed as follows. Stack F on top of F^{T} and compute the SVD of this $2n \times n$ matrix:

$$\begin{pmatrix} \mathsf{F} \\ \mathsf{F}^{\mathsf{T}} \end{pmatrix} = \mathsf{U}\,\Sigma\,\mathsf{V}^{\mathsf{T}}.$$

Using similar arguments as above, the matrix on the left is of rank 3 in general and the three rows of V^{T} corresponding to non-zero singular values in Σ can directly be chosen as rows of B, which concludes this self-calibration algorithm.

A similar procedure was proposed by Claus and Fitzgibbon for the case of 3×6 back-projection matrices [98]. The self-calibration approach for para-catadioptric cameras by Geyer and Daniilidis (cf. Section 5.3.1) follows the same principle [166, 170]. In their case, the back-projection matrix could be fully determined (up to a rotation instead of a projective transformation), due to using constraints on the coefficients of B. More generally speaking, there are two ways of obtaining more than the self-calibration of "just" the non-perspective part of B. (i) Use constraints on the entries of B. (ii) In any case, once B is estimated up to a projective transformation, we are in the situation of an uncalibrated perspective camera. From the fundamental matrix between two perspective images, one may estimate up to two intrinsic parameters [206, 474]; usually one just computes the focal length. Hence, in many cases it should be possible to obtain a full self-calibration of non-perspective cameras from a single fundamental matrix.

From the above, it is straightforward to extend self-calibration to the use of multiple fundamental matrices for a better stability and/or to the self-calibration of two or more cameras with different parameters/models. Without giving a formal proof, it should be possible for example, to self-calibrate three entirely different cameras (possibly with different sizes of back-projection matrices), from the three pairwise fundamental matrices. This possibility was demonstrated by Geyer and Daniilidis for para-catadioptric cameras [166, 170].

In the following, we review a few self-calibration approaches proposed in the literature for different types of non-perspective camera models.

Classical distortion model. Zhang proposed an approach for the simultaneous estimation of one radial distortion coefficient and the perspective fundamental matrix, by non-linear optimization of the epipolar constraint [566]. Based on his experimental results, Zhang concluded that the distortion coefficients can only be estimated reliably if matching points can be extracted accurately *and* if the actual distortion is significant (cf. the discussion in Section 3.6). In other cases, the estimation is ill-conditioned or in other words, taking into account distortion coefficients besides the fundamental matrix constitutes an overparameterization. One may further add that it seems to be required that image matches be well distributed across the image planes, as is always required when estimating radial distortion, which is most observable toward the borders of images.

In [461], Stein proposed a virtually identical method, as well as an extension to three views. Here, trifocal transfer errors are used to define a cost function which is optimized simultaneously over the distortion and trifocal tensor coefficients.

Non-central panoramas. Huang et al. proposed a self-calibration method for non-central panoramic images acquired by rotating 1D cameras [239]. In their approach, it is required that images be obtained by rotating about the same axis.

Catadioptric cameras. Gluckman and Nayar showed how to determine the focal length of the camera in a single-lens catadioptric stereo

system using one camera and two planar mirrors [174, 175]. Since this setup is identical to a system of two perspective cameras, perspective self-calibration approaches can be readily applied. Gluckman and Nayar used point matches between the two views and the fundamental matrix estimated from them. Their reported results were not perfect. We believe that this is because of a generic degeneracy: whenever the two mirrors are arranged in a rotationally symmetric position with respect to the camera's optical axis, the two virtual cameras (the real camera reflected in the two mirrors) are in a degenerate relative pose for focal length computation [474]. In situations close to this, results may be expected to be inaccurate.

The first self-calibration approach for non-perspective catadioptric systems is probably due to Kang, who considered the central para-catadioptric case [262]. Based on an expression of the epipolar geometry as a function of the relative motion between two views and the relevant intrinsic parameters, he proposed a cost function consisting of the sum of distances between image points and epipolar circles. This cost function was minimized non-linearly.

Geyer and Daniilidis showed how to self-calibrate para-catadioptric cameras from point matches, via the estimation of the 4×4 fundamental matrix [166, 170]. Their key observation was that the null-space of the fundamental matrix contains the vector representation of the image of the absolute conic (a circle in their case). They showed that the fundamental matrix is of rank 2; if both images are taken by the same camera, then the left and right null-spaces may be intersected to find the image of the absolute conic and thus to calibrate the camera. Geyer and Daniilidis also showed that if the images are taken with different para-catadioptric cameras, then their intrinsic parameters can be found from the fundamental matrices between three images. These observations can be generalized, cf. the above paragraph "General observations".

Sturm showed that a similar concept can be used for the self-calibration of para-catadioptric cameras from homographies associated with scene planes [471].

In [344, 347, 349], Mičušik and Pajdla proposed an approach for the estimation of the epipolar geometry of non-central catadioptric

cameras, demonstrated for the examples of a spherical mirror (always non-central) and misaligned para- and hyper-catadioptric systems. The approach proceeds in several steps. First, an appropriate central approximation to the full non-central back-projection is derived. This approximate model is then linearized such that its coefficient(s) can be estimated simultaneously with the fundamental matrix, by solving a quadratic eigenvalue problem similarly as in Fitzgibbon's approach for the division model (see below). This allows to initialize and then optimize the complete, non-central model.

Polynomial and rational polynomial models. Fitzgibbon showed how to simultaneously estimate radial distortion and perspective multi-view relations such as the fundamental matrix, plane homographies, or trifocal tensors, by solving quadratic or cubic eigenvalue problems [145]. The approach is based on the division model (see Section 3.1.8). This was extended by Mičušik and Pajdla [344, 346, 349] (see previous paragraph), Barreto and Daniilidis [33, 34], Claus and Fitzgibbon [98], and Steele and Jaynes [458].

Claus and Fitzgibbon showed how to estimate the coefficients of the full bi-quadratic rational function model for central cameras (i.e., 3×6 back-projection matrices) from point matches between two views [98]. They also analyzed the nature of epipolar curves (conics) and showed how to calibrate the model from one image of a planar calibration grid.

Mičušik and Pajdla used a division-type model as an extension to the equiangular fisheye model (see Section 3.1.7) and showed, analogously to Fitzgibbon's approach, how to simultaneously estimate its two parameters and the fundamental matrix [344, 346, 349]. If both parameters of their model are used, 15 point matches are required; for a one-parameter version 9 matches are sufficient. They used an analog approach to self-calibrate a one-parameter model of para-catadioptric cameras, leading to a quartic instead of a quadratic eigenvalue problem [344, 348, 349].

Barreto and Daniilidis showed that even if two images with different distortion coefficients for the division model are considered, they are subject to a 4×4 fundamental matrix [34]. They gave an algorithm for linearly estimating the distortion coefficients and the epipolar

geometry, from a minimum of 15 point matches. The same factorization approach is used in [33] to calibrate the nodes of a multi-camera network, including radial distortion, intrinsic and extrinsic parameters.

Li and Hartley proposed a different method for estimating distortion coefficients of the division model from point matches in two views [311, 312]. They decoupled the estimation of distortion coefficients from that of the fundamental matrix and showed how to estimate one or two coefficients of the division model from 9 point matches.

Kúkelová and Pajdla proposed a minimal method for the self-calibration problem with one distortion coefficient, using 8 point matches [292]. In [72, 291, 293], Kúkelová, Byröd et al. gave minimal methods for the following two problems. (i) Estimating the fundamental matrix together with two different coefficients for the division model, from 9 point matches (a non-minimal linear solution for this problem was first given by Barreto and Daniilidis [34], see above). (ii) Estimating the epipolar geometry and one distortion coefficient (identical for both images) in the case of partially calibrated images (known principal point and square pixels); a minimal method for estimating the distortion coefficient and the essential matrix (as opposed to the fundamental matrix) was given.

Steele and Jaynes extended Fitzgibbon's approach in two ways [458]. First, they solved a more general quadratic eigenvalue problem, allowing to obtain less ambiguous and more accurate results in the case of more than the minimum of 9 point matches. Second, their approach may be used to simultaneously estimate a global distortion coefficient and fundamental matrices for multiple image pairs, thus increasing the accuracy of each of these estimates.

Kannala et al. proposed a bundle adjustment type approach for the self-calibration of a unified central catadioptric model (Section 3.1.5) and the polynomial model of Equation (3.7), from matches between two images [267].

Discrete camera models. Thirthala and Pollefeys showed how to self-calibrate the 1D radial camera model (see Section 3.1.9) from four images, based on estimating the associated quadrifocal tensor [500]. Their approach is stratified, starting with a projective reconstruction

and converting it to a Euclidean one. At that stage, the radial distortions of the camera are not yet calibrated, just the mapping from radial planes to radial lines. To get a full calibration, Thirthala and Pollefeys used the assumption of radial symmetry. They either fit a division model to the reconstruction [501] or a discrete camera model.

In addition to their approaches for grid-based calibration of general radial distortion models (see Section 5.1.2), Hartley and Kang also showed that self-calibration is possible [209]. They observed however that results are too unstable in the presence of even small amounts of noise.

Tardif et al. proposed an approach for the self-calibration of a general radially symmetric radial distortion model (cf. Section 3.3.2) from images of a planar scene with unknown structure [489]. A more robust extension which also works for images with different radial distortions was presented in [490]. A factorization-based method for the same problem was given by Ramalingam et al. [417].

5.3.2 Special or Known Motions

Like for perspective cameras, using special types of camera motion or actually knowing the camera motion usually allows to simplify the self-calibration problem. This enables less complex algebraic problem formulations, allows to work with fewer images, and often gives more accurate results. As for perspective cameras [207], the most used motion is pure rotation about the optical center; an approximate rotation is sufficient if the scene is sufficiently far away. In the following, we describe several approaches, first ones based on special types of motion (but with unknown motion parameters), then ones using knowledge of the motion parameters.

Classical distortion model. Stein proposed a method based on pure rotational camera motion [460]. Matching (undistorted) image points are related by a special homography. Distortion coefficients and that homography are estimated simultaneously, by non-linear optimization, using a reprojection error type cost function.

Sawhney and Kumar showed the importance of estimating non-perspective distortion for generating mosaics based on images

acquired with a rotating camera [435]. They proposed a direct method (i.e., based on image intensities rather than feature matches) for simultaneously estimating distortions and image alignment transformations. In their work, they only considered a single term of the classical radial distortion model, but it is obvious that other distortion models may be estimated from images taken under pure rotation.

Fish-eye models. Xiong and Turkowski performed the self-calibration of a fisheye camera from pure rotations [539]. The camera model they used is the extension of the equiangular model described in Section 3.1.7. The calibration is performed by non-linear minimization, where the distortion center/principal point is initialized as the center of the image's bounding ellipse.

Discrete camera models. Thirthala and Pollefeys showed how to self-calibrate the 1D radial camera model (see Section 3.1.9) from three images taken under pure rotational motion [500, 501]. To be precise, in the non-central version of the 1D radial model, pure rotation would rather correspond to camera poses such that all optical axes intersect in a single point. A full self-calibration algorithm was given for central cameras.

Tardif et al. showed how to self-calibrate a general radially symmetric radial distortion model using either pure rotational or pure translational camera motion [489]. They used both a parametric radial distortion curve and a discrete sampling thereof.

The probably most general self-calibration problem considered to date, besides the work of Grossmann et al. (see Section 5.3.3), concerns the ray-based self-calibration of a general central camera. In the following works, the only assumption made is that of continuity — neighboring image points have neighboring camera rays.

Ramalingam et al. showed that from pure translational motions, the self-calibration of camera rays is possible up to a projective transformation of their directions [419]. They stated that a complete self-calibration is possible from two rotational motions and gave an algorithm for full self-calibration from one translational and two rotational motions. In [421], they proposed an algorithm that only requires two rotational motions.

Nistér et al. and Grossmann et al. [193] gave several similar results, mainly for the case of infinitesimal rotations or translations [380, 193]. Their main result is an algorithm for self-calibration from rotational optical flow. They issued the conjecture that two infinitesimal rotations about different axes are sufficient for a complete self-calibration. This was theoretically proven by Espuny and Burgos Gil who also gave a practical algorithm [130, 131].

Known motions. Stevenson and Fleck developed an approach using a robot arm and a light source to calibrate a camera [462]. The camera is mounted on the robot arm which then rotates the camera by known angles (the rotation should be about the camera's optical center, although this is not further discussed in [462]). The image of the fixed light source is extracted in each frame. The approach calibrates a radially symmetric central model, as well as optical parameters such as light fall-off across the image plane. The gist of the proposed approach is as follows. An initial estimate of the distortion center is assumed; the camera is rotated such that the light source is seen at the distortion center. Then, the camera is successively rotated in one direction by some small angle increment. The observed image positions of the light source, together with the known rotation angles, directly allow to construct a discrete sampling of the distortion function. Stevenson and Fleck also described how to correct the initial estimate of the distortion center.

Similar active approaches were also proposed by Du and Brady [122] and Kojima et al. [282] and seem to be in use for the calibration of cameras for space and planetary exploration, see e.g., [334].

Swaminathan et al. proposed an approach for calibrating non-central catadioptric cameras, using image matches and knowledge of the camera motion [484]. The approach is based on a parametric model for the caustic of catadioptric cameras; its coefficients are non-linearly optimized in a bundle adjustment like manner.

5.3.3 Image Statistics

Farid and Popescu showed in [135] that distortion coefficients, as well as coefficients of non-linear intensity transformations, may be estimated

from single images without the use of calibration objects. This is based on a frequency analysis of the image and on the assumption that non-linear transformations (geometric or intensity distortions) introduce specific higher-order correlations in the frequency domain. In some sense, this approach is an example of using natural image statistics, here for camera calibration. A similar approach was developed by Yu [561].

Grossmann et al. showed that a highly discrete camera model, consisting of "isolated" camera rays (see Section 3.3.3), can be self-calibrated just by taking images of unknown scenes, based on simple assumptions about their appearance [192, 194]. Roughly speaking, it is assumed that the smaller the angle between two camera rays, the more likely it is that the associated pixels produce similar greylevels; this is in some sense, an assumption on the statistics of natural images. When acquiring a sufficiently large set of images, observed differences between greylevels of different pixels, allow to infer the topology of the camera rays. Based on a simple off-line training, one may even recover the angles between rays and finally, their directions in a common coordinate system.

5.4 Special Approaches Dedicated to Catadioptric Systems

In this section we only cover methods that were not explained in any of the above sections. When using catadioptric cameras, one rarely does not know anything at all about the shape of the mirror, unless it is taken "off-the-shelf" or if the producer of a catadioptric camera does not disclose any information about the mirror's shape. In case the mirror shape is known, calibration can be considered as estimating the relative pose between camera and mirror, and possibly the camera's intrinsic parameters. Otherwise, several possibilities exist, as explained below.

Known mirror shape. One of the two (self-)calibration methods proposed in [262] by Kang is based on the extraction of the mirror's upper circular rim in the image. From this, some or all intrinsic parameters can be computed. This idea was used in several other works, for example [110].

Strelow et al. proposed a calibration method using 3D calibration grids [467]. They assumed that the perspective camera looking at the mirror is already calibrated. The calibration of the catadioptric sensor then comes down to estimating relative mirror-to-camera pose. Strelow et al. estimated this together with the pose of the calibration grid, by non-linearly optimizing the reprojection error of calibration points. Lhuillier showed how to optimize the mirror-to-camera pose and the 3D position of points observed and matches in several images, via bundle adjustment [308].

Aliaga used a parabolic mirror, looked at by a perspective camera [12]. Hence, the catadioptric system is non-central. A calibration procedure was proposed that estimates the focal length of the perspective camera and its position relative to the mirror (the orientation seems to be assumed to be perfect, i.e., the optical axis to be parallel to the mirror axis), as well as a radial distortion coefficient.

Similar but more restrictive approaches exist that require knowledge of the mirror shape and the camera's intrinsic parameters, i.e., calibration comes down to estimating the relative mirror-to-camera pose. Mashita et al. [332] supposed that the perspective camera is calibrated and that the mirror's circular upper boundary can be extracted in the image. Then, classical algorithms for estimating the pose of a circle in 3D can be applied, e.g., [114], after which the pose of the entire mirror, relative to the camera, is known. This effectively gives the calibration of the entire catadioptric system. This is similar to the approach by Kang described above, which was developed for central para-catadioptric cameras, whereas the one by Mashita et al. is for perspective cameras and central as well as non-central setups.

Another such method was proposed by Fabrizio et al. where besides the upper boundary of the mirror, a second circular section of it must be observable in the image, e.g., the footprint of a black needle glued on the mirror of some prototypes of catadioptric systems [133]. From the images of these two circles, the mirror pose is computed along with some intrinsic camera parameters.

Unknown mirror shape. Several possibilities exist to perform calibration in this case, some of which are explained in the following.

If it is known that the catadioptric camera is one of the central types (cf. Section 2.3.1), then one may for example use any of the unified models (cf. Section 3.1.5) to calibrate it.

A more general approach was proposed by Caglioti et al. for the case where all that is assumed is that the mirror is a surface of revolution [78]. The perspective camera looking at the mirror can be positioned arbitrarily and the mirror shape can be arbitrary (besides being a surface of revolution), hence the system is non-central in general. The proposed calibration approach requires to extract the silhouette of the mirror in the image as well as the image of at least one straight line in the scene. This input allows to calibrate intrinsic camera parameters (focal length and principal point), the shape of the mirror (its profile) and the relative pose of mirror and camera.

It is of course possible to cut the calibration process in two steps, where in a first step one reconstructs the mirror shape using any of the many methods for specular surface reconstruction available in the literature, most of which are based on analyzing images of patterns reflected in the mirror (a representative list of references is out of scope for this manuscript). An alternative to these classical approaches is to use polarization imaging, as proposed for example by Morel et al. [352].

Naturally, another possibility is to calibrate the catadioptric system using any of the global, local, or discrete models described in the previous section. If the mirror is of arbitrary shape, only local or discrete models might work.

6

Structure-from-Motion

The goal of this section is to briefly describe the main building blocks of structure-from-motion in a simple way, that is independent of the camera model used. This is to highlight the underlying (simple) principles, allowing to establish links between seemingly different approaches that abound in the literature and to easily adapt these principles to new camera models.

The first three sections are dedicated to the main building blocks for *calibrated* cameras: pose and motion estimation, and triangulation of points. These will be explained for central and non-central cameras. The fourth section aims at highlighting the central role of bundle adjustment in all of structure-from-motion and calibration. The fifth section lists a few works on 3D scene modeling from omnidirectional cameras, from single or multiple views. In the last section, the related issues of distortion correction and stereo rectification are discussed; they are not really structure-from-motion topics, but necessary for 3D scene modeling.

Overall, we believe that the most important ingredients of structure-from-motion systems are: algorithms working with the minimum required amount of data, to be embedded in robust estimation schemes such as RANSAC, and a (robust) bundle adjustment, allowing to get

optimal results for all of the individual problems. Other important contributions, not discussed further in this section, are methods determining globally optimal solutions of structure-from-motion problems, see e.g., [260] and references therein. Besides the minimal methods, other algorithms that give suboptimal results may be useful, e.g., to get a better initial estimate from the set of inliers of a model determined using a minimal method. Finally, this section only covers some main structure-from-motion problems; many special cases exist, using lines or other primitives than the points considered in the following, concerning dynamic scenes or objects, etc.

6.1 Pose Estimation

Computing the pose relative to a calibrated perspective camera, of an object with known structure, is one of the most classical problems in photogrammetry (where it is called "resection") and computer vision, see, e.g., a review paper by Haralick et al. [205]. Pose, i.e., object position and orientation, is usually computed using point or line correspondences between the object and the image, although other possibilities exist, for example computing the pose of a spherical object from its silhouette in the image [114]. The minimum number of required point correspondences is three, in which case there may be up to four possible solutions [230]. Any of the numerous algorithms that were proposed for the perspective camera model can be easily adapted to any other central camera model, since instead of working with the original 2D image coordinates, each algorithm uses, explicitly or implicitly, the directions of the back-projected camera rays. Hence, these algorithms can be used for any central camera as soon as it is calibrated. There exist different formulations for the pose estimation problem. Here, we use one that is directly suited for central as well as non-central cameras. Minimal solutions for the non-central pose estimation problem were proposed independently by Chen and Chang [88], Nistér [378], Nistér and Stewénius [379], and Ramalingam and Sturm [414, 415]. See also papers by Fabrizio and Devars [132] and Schweighofer and Pinz [442] about the n-point pose problem for central and non-central cameras.

General problem formulation. The problem can be formulated as follows. Instead of explicitly estimating the rotation and translation of the object model, we estimate the depth of object points relative to the camera, i.e., their position along the known camera rays, associated with the matched image points. Given the position of 3 points, the rotation and translation for the entire object can then be computed easily (via absolute pose estimation, see for example [233]). We use the finite and infinite parts of back-projection, cf. the introduction of Section 3: a camera ray is defined by a finite point $\mathbf{B}^f$ (for central cameras, the optical center) and the direction $\mathbf{B}^i$. A point on the ray is then parameterized by a scalar λ:

$$\mathbf{Q} \sim \mathbf{B}^f + \lambda \begin{pmatrix} \mathbf{B}^i \\ 0 \end{pmatrix}.$$

With three object points $\mathbf{Q}_j$, parameterized by $\lambda_j, j = 1 \cdots 3$, we can set up three constraint equations: the Euclidean distances between these points must be identical to the prior known values. The squared distance between two points $\mathbf{Q}_j$ and $\mathbf{Q}_k$ is quadratic in λ_j and λ_k. From the three pairs of points, we thus get a total of three quadratic equations in three unknowns. One may use any appropriate numerical method for solving such a system; it is simple for example to convert the equation system into a degree-8 polynomial in one of the unknowns and then to use any numerical root finding method to get the up to eight real solutions. This should naturally be embedded in a robust estimation scheme, such as RANSAC, where each of the potential solutions is scored against the other point matches.

Central cameras. The usual special case of central cameras allows for a slight simplification. As finite point $\mathbf{B}^f$, the origin may be adopted for all camera rays; our points can then be parameterized as:

$$\mathbf{Q}_j \sim \begin{pmatrix} \lambda_j \mathbf{B}_j^i \\ 1 \end{pmatrix}.$$

If we swap the sign for all three λ_j, the distances between the $\mathbf{Q}_j$ are not affected. Hence, for each solution of our three constraint equations, there exists a mirror solution with all signs swapped. This allows to

reduce the complexity of the estimation problem: instead of solving the equivalent of a degree-8 polynomial, a degree-4 polynomial is sufficient.

Other pose problems. Josephson and Byröd gave a minimal method for estimating camera pose, together with the focal length and one radial distortion coefficient (division model), from four matches [259].

A special case is that of planar pose, where the unknown object pose consists of a translation in one plane (usually, a horizontal ground plane) and a rotation about the normal of that plane. Minimal solutions requiring 2 point matches only exist [12, 393]. Gourichon et al. proposed a bio-inspired method for planar pose estimation [184].

6.2 Motion Estimation

Generalities. Motion or egomotion estimation, or relative orientation, is another classical problem. The goal is to estimate the motion of a camera or the relative pose of calibrated cameras, from matches between images but without using information on the scene structure.

The first truly minimal solution for central cameras was developed by Nistér [377]: an approach for estimating all possible essential matrices [324] from 5 point matches (up to 10 theoretical solutions).

As for non-central cameras, the concept of essential matrix was introduced by Pless [402], cf. Section 4.1. It can be estimated by solving a linear equation system, established from 17 or more point matches; extracting the motion parameters from the essential matrix is then relatively straightforward. Note that if the cameras are not "fully non-central", e.g., are of the axial type, then such a linear estimation process may be underconstrained, as pointed out by Mouragnon et al. [357]. Other degeneracies were studied by Kim and Kanade [279]. Kim et al. proposed two motion estimation algorithms for the case where the non-central system consists of two or more central cameras with non-overlapping fields of view, i.e., where points are only tracked by individual cameras but where no handover between cameras occurs [277]. Finally, a minimal solution for the general non-central case, from 6 point matches, was given by Stewénius et al. [464].

Contrary to motion estimation from central cameras, the scale of the translation can also be estimated with non-central cameras. However, this is only accurate if the trade-off of several factors is favorable (cf. the general discussion in Section 3.6): the cameras should be sufficiently non-central, compared to the amount of noise, the distance of the scene from the cameras, etc. The other way round: if a camera is not sufficiently non-central, then it is advisable to model it as a central system and use traditional motion estimation, cf. also a study by Kim et al. [278].

In the following, we briefly cite a few works on motion estimation for omnidirectional cameras.

Works on motion estimation for omnidirectional cameras. Yagi et al. considered the motion estimation with a central hypercatadioptric camera [543]. The robot-mounted camera is supposed to move on a horizontal ground plane; the topic of [543] is to estimate small perturbances to perfect planar motion (so-called rolling), from optical flow.

McMillan and Bishop estimated the motion between two calibrated cylindrical images [338]. This is done by non-linear minimization of the distances between back-projection rays associated with matching image points. Kang and Szeliski [264] used the classical 8-point method for linearly estimating the essential matrix in the same case. Other approaches for motion estimation that are based on computing the essential matrix, explicitly or implicitly from directions of back-projected image points, are [86, 479]. Svoboda discussed the issue of properly normalizing coordinates of image matches to ensure a good conditioning [391, 479], along the lines of Hartley's approach [208]. Bazin et al. proposed a motion estimation method for catadioptric cameras that decouples the estimation of rotation and translation parameters [45]. The method is designed for man-made environments and requires the detection of sets of parallel lines.

Gluckmann and Nayar estimated motion from the optical flow [173]. In order to use a wide field of view and to use the same method for motion estimation for different imaging geometries, they mapped the flow field to a sphere centered in the optical center, using camera

calibration information. Then, extensions of classical algorithms for motion estimation from optical flow, from planar to spherical viewing surfaces, can be readily applied, independently of the actual camera's imaging geometry. We also refer to the seminal work on determining camera motion from the flow field on the viewing sphere, by Nelson and Aloimonos [373], where they also clearly explained why wide fields of view are superior to narrow ones for motion estimation. See also the papers [141] by Fermüller and Aloimonos who studied the same issue while drawing analogies with the evolution of biological eyes, [304] by Lee et al., and [375] by Neumann et al. who proposed a mathematical formulation for designing non-central cameras that are optimal for motion estimation. Shakernia et al. proposed a similar approach to that of Gluckman and Nayar, but instead of back-projecting optical flow to a sphere, they used another curved surface which is implicitly defined by the lifting operation carried out when back-projecting image points [446].

In [468], Strelow and Singh described how to combine image point matches in omnidirectional cameras with inertial sensor readings, for structure and motion estimation. A first approach assumes a calibrated camera. A second approach works for uncalibrated cameras — only the assumption of a radial projection function is required (cf. Section 3.1.9). This is highly related to work by Thirthala and Pollefeys, see Sections 3.1.9 and 5.3.1, and to the RAC (radial alignment constraint) by Tsai [515].

Shakernia et al. formulated motion estimation under the assumption of small displacements between images [447].

Lim et al. introduced a so-called antipodal-epipolar constraint that allows to decouple the estimation of rotation and translation for cameras with larger than hemispheric fields of view [318]. This result is based on the notion of antipodal image points: points whose camera rays are aligned but which correspond to opposite viewing directions. If two antipodal points in one camera can both be matched in the other camera, then a specific epipolar constraint becomes available. Lim et al. showed how to independently estimate translation and/or rotation from two or more pairs of such antipodal point matches. In [317], they showed

a similar decoupling approach whose inputs are optical flows measured at antipodal point pairs.

6.3 Triangulation

Points. Reconstructing a 3D point from matches in two or more calibrated images (called intersection in photogrammetry) is of course an essential module in structure-from-motion. Like for the other problems, an optimal solution requires in general a bundle adjustment or a global optimization approach. For the special cases of two or three perspective images, "direct" optimal solutions were proposed by Hartley and Sturm [211] and Stewénius et al. respectively [465].

Suboptimal but generic solutions are to estimate the 3D point that minimizes some distance measure relative to the camera rays associated with the input image matches, for example the Euclidean distance [478] or an angle-based measure, see Section 6.4.

Lines. The triangulation of lines or line segments is useful, especially in 3D modeling of man-made objects. Whereas for points, it does not matter if they are reconstructed from central or non-central cameras (a single ray is used per camera), there is a difference in the case of lines. As for central cameras, whatever type they are, a line image can be back-projected to an *interpretation plane* in 3D — the plane spanned by the optical center and the original 3D line. From two images, the line can thus be triangulated by computing the intersection of the two interpretation planes.

As for non-central cameras, they allow in general the triangulation of 3D lines from single images. The reason is that the back-projection is no longer a plane but some other surface and in general, the original 3D line is the only straight line on that surface. Another way of formulating this is by considering image points lying on a line image, and the associated camera rays. It is well known that for four lines in general position, there exist two lines incident to all of them. If more than four camera rays associated with a line image are considered, the original 3D line can be found as the only one incident to all of them, using the approach of Teller and Hohmeyer [494]. Of course, if the camera

is of the axial type, all camera rays are incident with the camera axis; this can be taken care off by computing all candidate 3D lines (two in general) and by discarding the camera axis. There may be ambiguous solutions for the 3D line, depending on the camera used and the position of the original line.

Algorithms and in-depth theoretical studies for line triangulation from non-central images are given by Caglioti and Gasparini [75, 76, 77] as well as Lanman et al. [299]. Morita et al. described an early method for reconstructing 3D lines from two fisheye images [353]. Although not described as such, the method seems equivalent to intersecting interpretation planes. Like other researchers, for example Vasseur and Mouaddib [519], Morita et al. used the fact that 3D lines in spherical images, appear as great circles. Yamazawa et al. described a Hough-transform-based approach for detecting line segments in omnidirectional images and demonstrated the 3D reconstruction of lines from image triplets [545]. Another approach for 3D line (segment) triangulation for omnidirectional images is for example [242].

6.4 Bundle Adjustment

As mentioned earlier, bundle adjustment is imperative for obtaining an accurate solution of any structure-from-motion, calibration, or other similar problem, an exception being problems for which global optimizers exist. Bundle adjustment means in principle nothing else than a simultaneous non-linear optimization of all unknowns; in addition, it implies that a meaningful cost function is used, see below. Other important issues are the exploitation of the problem structure at hand to obtain efficient algorithms; this includes the well-known sparsity of the normal equations solved at each iteration of the usually employed non-linear least-squares methods but also the question of how to order data in the case of image sequences with multiple overlaps. More on bundle adjustment can be found in the overview paper by Triggs et al. [512].

Bundle adjustment is most often a non-real-time process. Recently, approaches that may be qualified as hybrids between bundle adjustment and simultaneous localization and mapping (SLAM) [107] were proposed, by Mouragnon et al. [356, 358] and Engels et al. [129]. SLAM,

mostly used in robotics, solves in principle exactly the same problem as structure-from-motion: estimating camera motion and scene structure. The main difference is the real-time requirement in robotics applications, which lead to sequential solutions for cameras and other sensors (SLAM), whereas bundle adjustment and structure-from-motion systems were often developed without having to satisfy this requirement. The hybrid approaches of Mouragnon et al. and Engels et al. allow for real-time performance by following an overall sequential scheme, while continuously performing a complete non-linear optimization on as long subsequences as possible with the available computing power.

As mentioned above, an important aspect of bundle adjustment is the choice of cost function. As generally agreed, the "best" cost function is based on the reprojection error, i.e., the distance between measured image points and those predicted by the unknowns (scene structure, camera intrinsics or extrinsics). To be more precise, the cost function should be defined based on a statistical model of the noise in the input data, whenever this is available. If the input, as usual, are extracted image points then one should use an uncertainty estimate of the interest point extractor; under the default assumption of i.i.d. Gaussian noise this automatically leads to the standard cost function defined as the sum of squared distances between measured and predicted image points. Ideally, actual uncertainty estimates of image point positions should be used.

Minimizing reprojection errors requires carrying out forward projections of 3D points; as described elsewhere in this manuscript, for some cameras back-projection is easier to formulate. Further, if already calibrated cameras are considered, then one may develop a generic bundle adjustment working with camera rays instead of image points as basic input. Such procedures were for example proposed by Ramalingam et al. [414, 415], Schweighofer et al. [443], and Lhuillier [309, 310].

Ramalingam et al. used as cost function, the distance between camera rays and 3D points, which is simple to use but suboptimal. Schweighofer et al. gave a provably convergent algorithm for estimating camera motion and 3D structure, using the same cost function [443]. As an alternative solution for non-central cameras, Ramalingam et al. proposed to partition the set of camera rays into

clusters of approximately convergent rays, which are then treated as individual pinhole cameras inside a classical reprojection error-based bundle adjustment. Such an idea was also suggested though apparently not implemented, by Martins et al. [331].

Lhuillier proposed and justified the use of an angle-based cost function [309, 310]. To be precise, he suggested to use the squared tangent of the angle between the back-projected camera ray (computed from estimated camera parameters) and the line spanned by the optical center and the estimated 3D point. He also discussed how to minimize the reprojection error for catadioptric cameras.

Kannala et al. used a different angle-based cost function for bundle adjustment (sum of squared sines) [267], inspired by Oliensis [385].

6.5 Three-Dimensional Scene Modeling

In this section, we cite a few approaches for complete 3D scene modeling from omnidirectional images. Most of them rely on a combination of the above structure-from-motion techniques and binocular or multi-view dense stereo methods.

Kato et al. used a shape-from-silhouette approach with a fisheye camera to reconstruct a moving object, whose motion was supposed to be determinable [270].

Kang and Szeliski used mosaics acquired by a calibrated perspective camera and expressed in cylindrical coordinates [264]. After motion estimation and bundle adjustment based on interest points, they carried out a dense multi-baseline stereo algorithm. A similar system, active-vision based, was proposed by Ishiguro et al. [247]. Teller and his group used georeferenced omnidirectional images for city modeling [493]. Bunschoten and Kröse [68, 69], Doubek and Svoboda [119], Mičušik et al. [345], and Havlena et al. [214] presented other multi-baseline stereo approaches for central omnidirectional images.

One of the most advanced approaches, demonstrated on long image sequences, is that of Lhuillier [310].

There also exist approaches for 3D modeling from single omnidirectional images. Sturm proposed a method for para-catadioptric cameras, exploiting user-provided geometric scene constraints such as

perpendicularity or parallelism of lines and coplanarity of points [470]. The method consists of two steps, first a calibration step, then the actual 3D modeling. It is important to underline that the second step is not specific to omnidirectional images, but could be applied to any calibrated central image: the available geometric constraints are applied on the back-projected camera rays and it does not matter what the underlying imaging geometry is.

Another single-view 3D modeling approach for omnidirectional images is due to Chen and Ip [93]. The same comment as above applies, i.e., the actual 3D modeling method is valid for any central camera and not specific to omnidirectional ones. Chen and Ip used a rectangular object in fronto-parallel position, to calibrate a radial lens distortion model and used that object also as reference plane for the 3D model.

6.6 Distortion Correction and Rectification

By distortion correction we understand here the generation of perspectively correct images from distorted ones. This is highly related to the problems of texture mapping and stereo rectification. Let us first consider texture mapping from one image: given a calibrated image and some surface, the goal is to "paint" the surface by back-projecting the image and its intensities onto it. Geometrically, this could be done by considering one pixel after the other; for each pixel, we determine the intersection of its camera ray and the surface and apply the pixel's intensities at that point. It is well known that such a process may lead to holes in the paint and other artifacts and that it is recommended to proceed in the opposite direction. To do so, let us consider a representation of the texture map as a 2D digital image. In addition, we need the mapping between positions in the texture map and points on the 3D surface. The texture map is then computed similar to the above but the other way round: for each pixel of it, we compute the corresponding point on the surface and project it into the input image. The intensities applied to the texture map pixel are then computed using some interpolation scheme, from the intensities in the input image in the vicinity of the computed image position. This is a standard procedure and can be applied to any type of surface and camera.

The application of this procedure to distortion correction is straightforward: it suffices to choose a plane as 3D surface for this texture mapping. The result will be a perspectively correct image, whenever the input image was taken by a central camera (see Figure 6.1). For non-central cameras though, generation of perspectively correct images is only possible in general if the scene geometry is known (one can then texture map the scene and render it from any desired viewpoint with a

Fig. 6.1 Examples for distortion correction. Left and middle: A fisheye image and distortion corrected sections of it. Right: A cubemap generated from a fisheye image.

virtual perspective camera). Swaminathan et al. proposed an approach to obtain good approximations of perspective views if the scene structure is unknown but if simple priors on it are available [483]. A similar approach, for misaligned hyper-catadioptric cameras, was proposed by Jeng and Tsai [256]. Ding and Yu proposed an interactive approach for minimizing distortions in multi-perspective images that can be modeled using the general linear camera model (GLC, see Section 3.4.4) [116]. This problem is an instance of the more general image-based rendering problem [181, 307].

As a sidenote, let us mention an optical procedure for generating perspective images from a fisheye image, explained and demonstrated in 1925 by Beck [48]: he placed the negative of the fisheye image in the camera and reversed the action of the latter, by illuminating the negative and placing a photoplate in front of the camera. The photoplate then gets imprinted with a perspective photograph corresponding to the part of the fisheye's field of view covered by the plate.

Using planar 3D surfaces for distortion correction has its limits, since a hemispheric field of view would require an infinitely large texture map to hold the entire distortion-corrected input image. An alternative is to use cubemaps, i.e., where partial perspective view are rendered onto the insides of a virtual cube, see Figure 6.1 and for example the paper [265] by Kangni and Laganière. Cubemaps and other useful projections for environment mapping were well explained by Greene [187]. See also the work [564] by Zelnik-Manor et al. for various other suggestions for projections used to warp spherical images to flat ones.

The above texture mapping scheme can be applied for other tasks than the correction of non-perspective distortions. A typical example is to warp the usual circular omnidirectional images, which are hard to interpret by humans, into panoramic images; in that case, the 3D surface is the inside of a cylinder and the texture map corresponds to the unrolled cylinder. Besides this, the generation of the texture map is done exactly the same way as above.

Another application is stereo rectification. This consists in warping an image pair such that the warped images have an epipolar geometry where epipolar curves correspond to scanlines, i.e., where they are straight lines and such that in addition corresponding epipolar lines

have the same vertical image coordinate. This enables efficient implementations of stereo matching algorithms. As for perspective images, rectification can be achieved by applying appropriate homographies to each image (there exist infinitely many solutions for these homographies; different methods differ in the choice of a particular solution), see for example [21, 65, 176, 212, 325, 451].

Geyer and Daniilidis proposed a stereo rectification method for para-catadioptric cameras, which applies conformal warping transformations [169]. Banno and Kanade showed a rectification method for spherical images [24]. Heller and Pajdla proposed a stereographic rectification procedure for omnidirectional stereo pairs [217]. Instead of mapping epipolar curves to scanlines, they suggested to map them to circles; this is based on the observation that this way of rectifying minimizes, in some sense, the distortions between original and rectified images.

In 3D, one may picture stereo rectification as follows. Rectified images correspond to perspective cameras which are oriented identically (i.e., their optical axes are parallel and the image planes are rotated the same way) have the same focal length and whose baseline is parallel to the horizontal image direction. Given an input stereo pair, one may transform it into a rectified geometry, as follows. First, compute the baseline and choose any planar surface that is parallel to the baseline, at some finite distance from it. That surface will play the role of image plane for the rectified images. Next, for each camera, compute the rotation that would make the optical axis orthogonal to that plane and such that the image plane's horizontal axis becomes parallel to the baseline. Finally, one has to zoom in or out one or both of the views by appropriate factors that lead to equal focal lengths for both images. Rotating a camera in 3D and changing its focal length is equivalent to applying an appropriate projective transformation on its 2D image, the homography used in rectification algorithms.

This principle can be applied to any central camera, even non-perspective ones. The warping transformations are no longer homographies in general but this does not matter; all that is needed is the forward projection for the cameras, the determination of the rectified

image plane, and the application of the above distortion correction method.

One problem with this approach is that a planar image surface is not adequate for handling wide field of view cameras, as explained above. A second main problem is that the case where an epipole lies inside the image cannot be handled; it would again require an infinitely large rectified image.

For a fully general solution, one thus has to use other projection surfaces than planes. Roy et al. used a cylindrical surface, whose axis is orthogonal to the baseline [431]. The final rectified images are obtained by projecting the input images on the cylinder and unrolling it. If a parametric projection model is available for both cameras, then the mapping from rectified to original images can be written in closed form.

Fig. 6.2 Example of the rectification procedure of Pollefeys et al. On the left, two input images (the epipoles are marked as white dots), on the right the two rectified images. Courtesy of Marc Pollefeys [404].

Abraham and Förstner showed such a rectification procedure for fisheye cameras [1].

Pollefeys et al. proposed a yet more general approach, which is applicable even without a closed-form projection model and which exploits the fact that in true cameras, the search for matches can usually be restricted to half-epipolar lines [404]. Essentially, the approach generates rectified images by straightening individual epipolar curves, as follows (cf. Figure 6.2). The set of pairs of epipolar curves are browsed. For each pair, a new row is added to each rectified image. Roughly speaking, the intensities of pixels along each epipolar curve are just plotted one after the other on that new row of pixels. It does not matter if or if not the position of the plotted pixel in the rectified image is in any simple parametric relationship to the pixel in the input image; all that matters is that all intensities found along an epipolar curve are reproduced on a single row in the rectified image. This approach is highly general and is even applicable to non-central cameras, as long as they satisfy the standard stereo geometry, cf. Section 3.4.2. A conceptual difference to the above approaches is that in general, it may not correspond to a projection on (or, texture mapping of) some 3D surface.

7

Concluding Remarks

We have described some image acquisition technologies and associated camera models, with an emphasis on omnidirectional cameras. Principles and existing works on basic issues for geometric computer vision, such as epipolar geometry, calibration, and structure-from-motion, were reviewed. One of the goals in writing this monograph was to give a survey of the literature and to highlight links between existing works and underlying principles. While aiming at a certain level of exhaustiveness in the coverage of the literature, such a survey can never be complete; we apologize for any significant omission. The reader is invited to notify the authors of works that should be cited; an addendum to this monograph shall be maintained on the first author's website.

Let us briefly summarize what we consider some of the main issues we were trying to convey. We concentrated on models for back-projection, in order to stress that these enable a simple and systematic formulation of epipolar geometry, self-calibration, pose and motion estimation, and the projection of 3D lines. For distortion correction and bundle adjustment however, forward projection should ideally be used. An important issue we discussed concerns the choice of camera model, for example if or if not a non-central model is appropriate in

some application. Our conclusion is that there is no single and simple answer, but that this choice depends on multiple factors such as the "true" camera geometry, the amount of noise in data, the distribution of calibration points, and so forth.

Finally, we note that this monograph is restricted to geometric aspects of omnidirectional vision. Equally if not more important, are optical and photometric characteristics; these are well explained and studied for example by Okatani and Deguchi [384] and in various papers by Nayar et al.

Acknowledgments

P. Sturm is grateful to the Alexander-von-Humboldt Foundation for a Research Fellowship supporting his sabbatical at TU Munich, during which the largest part of this manuscript was written. He also acknowledges support by the French ANR project CAVIAR which supported part of the research carried out that eventually lead to the idea of writing this monograph. J. Barreto and P. Sturm are thankful for their support by the French-Portuguese collaboration program PHC Pessoa. The authors wish to thank Tomáš Pajdla for comments on the manuscript and Aseem Agarwala, Frédéric Devernay, Sujit Kuthirummal, Marc Pollefeys, Ryusuke Sagawa, and Steve Seitz for kindly allowing us to use images of their work.

References

[1] S. Abraham and W. Förstner, "Fish-eye-stereo calibration and epipolar rectification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 5, pp. 278–288, 2005.

[2] E. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 99–106, February 1992.

[3] G. Adorni, M. Mordonini, S. Cagnoni, and A. Sgorbissa, "Omnidirectional stereo systems for robot navigation," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[4] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski, "Photographing long scenes with multi-viewpoint panoramas," in *Proceedings of SIGGRAPH, Boston, USA*, pp. 853–861, 2006.

[5] A. Agrawal, Y. Taguchi, and S. Ramalingam, "Analytical forward projection for axial non-central dioptric & catadioptric cameras," in *Proceedings of the 11th European Conference on Computer Vision, Heraklion, Greece*, pp. 129–143, 2010.

[6] M. Agrawal and L. Davis, "Camera calibration using spheres: A semi-definite programming approach," in *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France*, pp. 782–791, 2003.

[7] M. Ahmed and A. Farag, "A neural approach to zoom-lens camera calibration from data with outliers," *Image and Vision Computing*, vol. 20, no. 9–10, pp. 619–630, 2002.

[8] M. Ahmed and A. Farag, "Nonmetric calibration of camera lens distortion: Differential methods and robust estimation," *IEEE Transactions on Image Processing*, vol. 14, pp. 1215–1230, August 2005.

[9] M. Ahmed, E. Hemayed, and A. Farag, "Neurocalibration: A neural network that can tell camera calibration parameters," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece*, pp. 463–468, 1999.

[10] O. Ait-Aider, N. Andreff, J.-M. Lavest, and P. Martinet, "Simultaneous object pose and velocity computation using a single view from a rolling shutter camera," in *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, (H. Bischof and A. Leonardis, eds.), pp. 56–68, Springer-Verlag, 2006.

[11] S. Al-Ajlouni and C. Fraser, "Zoom-dependent calibration for consumer grade-cameras," in *Proceedings of the ISPRS Commission V Symposium on Image Engineering and Vision Metrology, Dresden, Germany*, pp. 20–25, September 2006.

[12] D. Aliaga, "Accurate catadioptric calibration for real-time pose estimation in room-size environments," in *Proceedings of the 8th IEEE International Conference on Computer Vision, Vancouver, Canada*, pp. 127–134, 2001.

[13] J. Aloimonos, "Perspective approximations," *Image and Vision Computing*, vol. 8, pp. 179–192, August 1990.

[14] L. Alvarez, L. Gómez, and J. Sendra, "An algebraic approach to lens distortion by line rectification," *Journal of Mathematical Imaging and Vision*, vol. 35, pp. 36–50, September 2009.

[15] J. Arnspang, H. Nielsen, M. Christensen, and K. Henriksen, "Using mirror cameras for estimating depth," in *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns, Prague, Czech Republic*, pp. 711–716, 1995.

[16] N. Asada, A. Amano, and M. Baba, "Photometric calibration of zoom lens systems," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 186–190, IEEE Computer Society Press, August 1996.

[17] K. Asari, S. Kumar, and D. Radhakrishnan, "A new approach for nonlinear distortion correction in endoscopic images based on least squares estimation," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 345–354, April 1999.

[18] C. Aschenbrenner, "Neue Geräte und Methoden für die photogrammetrische Erschließung unerforschter Gebiete," *Bildmessung und Luftbildwesen*, vol. 4, no. 1, pp. 30–38, 1929.

[19] K. Atkinson, ed., *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, 1996.

[20] O. Avni, T. Baum, G. Katzir, and E. Rivlin, "Recovery of 3D animal motions using cameras and mirrors," *Machine Vision and Applications*, vol. 21, no. 6, pp. 879–888, 2010.

[21] N. Ayache, *Stereovision and Sensor Fusion*. Cambridge, MA, USA: The MIT Press, 1990.

[22] S. Baker and S. Nayar, "A theory of single-viewpoint catadioptric image formation," *International Journal of Computer Vision*, vol. 35, no. 2, pp. 1–22, 1999.

[23] H. Bakstein and T. Pajdla, "Panoramic mosaicing with a 180° field of view lens," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pp. 60–68, 2002.

[24] A. Banno and K. Ikeuchi, "Omnidirectional texturing based on robust 3D registration through Euclidean reconstruction from two spherical images," *Computer Vision and Image Understanding*, vol. 114, no. 4, pp. 491–499, 2010.

[25] J. Barreto, "General central projection systems: Modeling, calibration and visual servoing," PhD thesis, Department of Electrical and Computer Engineering, University of Coimbra, Portugal, September 2003.

[26] J. Barreto, "A unifying geometric representation for central projection systems," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 208–217, 2006.

[27] J. Barreto, "Unifying image plane liftings for central catadioptric and dioptric cameras," in *Imaging Beyond the Pinhole Camera*, (K. Daniilidis and R. Klette, eds.), Springer-Verlag, August 2006.

[28] J. Barreto and H. Araújo, "Issues on the geometry of central catadioptric image formation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 422–427, 2001.

[29] J. Barreto and H. Araújo, "Geometric properties of central catadioptric line images," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pp. 237–251, 2002.

[30] J. Barreto and H. Araújo, "Paracatadioptric camera calibration using lines," in *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France*, pp. 1359–1365, 2003.

[31] J. Barreto and H. Araújo, "Geometric properties of central catadioptric line images and their application in calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1327–1333, 2005.

[32] J. Barreto and H. Araújo, "Fitting conics to paracatadioptric projections of lines," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 151–165, 2006.

[33] J. Barreto and K. Daniilidis, "Wide area multiple camera calibration and estimation of radial distortion," in *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, 2004.

[34] J. Barreto and K. Daniilidis, "Fundamental matrix for cameras with radial distortion," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 625–632, 2005.

[35] J. Barreto and K. Daniilidis, "Epipolar geometry of central projection systems using veronese maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 1258–1265, 2006.

[36] J. Barreto, J. Roquette, P. Sturm, and F. Fonseca, "Automatic camera calibration applied to medical endoscopy," in *Proceedings of the 20th British Machine Vision Conference, London, England*, 2009.

[37] J. Barreto, R. Swaminathan, and J. Roquette, "Non parametric distortion correction in endoscopic medical images," in *Proceedings of 3DTV-CON Conference on The True Vision, Capture, Transmission and Display of 3D Video, Kos, Greece*, 2007.

[38] M. Barth and C. Barrows, "A fast panoramic imaging system and intelligent imaging technique for mobile robots," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Osaka, Japan*, pp. 626–633, 1996.

[39] O. Baschin, "Eine einfache Methode der stereophotogrammetrischen Küstenvermessung," *Petermanns Mitteilungen*, 1908.

[40] Y. Bastanlar, L. Puig, P. Sturm, J. Guerrero, and J. Barreto, "DLT-like calibration of central catadioptric cameras," in *Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France*, October 2008.

[41] A. Basu and S. Licardie, "Alternative models for fish-eye lenses," *Pattern Recognition Letters*, vol. 16, pp. 433–441, April 1995.

[42] J. Batista, H. Araújo, and A. de Almeida, "Iterative multi-step explicit camera calibration," in *Proceedings of the 6th IEEE International Conference on Computer Vision, Bombay, India*, pp. 709–714, January 1998.

[43] G. Batog, X. Goaoc, and J. Ponce, "Admissible linear map models of linear cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA*, 2010.

[44] E. Bayro-Corrochano and C. López-Franco, "Omnidirectional vision: Unified model using conformal geometry," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, pp. 536–548, 2004.

[45] J. Bazin, C. Demonceaux, P. Vasseur, and I. Kweon, "Motion estimation by decoupling rotation and translation in catadioptric vision," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 254–273, 2010.

[46] S. Beauchemin and R. Bajcsy, "Modelling and removing radial and tangential distortions in spherical lenses," in *Proceedings of the 10th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany*, (R. Klette, T. Huang, and G. Gimel'farb, eds.), pp. 1–21, Springer-Verlag, March 2000.

[47] S. Beauchemin, R. Bajcsy, and G. Givaty, "A unified procedure for calibrating intrinsic parameters of fish-eye lenses," in *Proceedings of Vision interface, Trois-Rivières, Canada*, pp. 272–279, 1999.

[48] C. Beck, "Apparatus to photograph the whole sky," *Journal of Scientific Instruments*, vol. 2, no. 4, pp. 135–139, 1925.

[49] R. Benosman, E. Deforas, and J. Devars, "A new catadioptric sensor for the panoramic vision of mobile robots," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, 2000.

[50] R. Benosman and S. Kang, "A brief historical perspective on panorama," in *Panoramic Vision: Sensors, Theory, and Applications*, (R. Benosman and S. Kang, eds.), pp. 5–20, Springer-Verlag, 2001.

[51] R. Benosman and S. Kang, eds., *Panoramic Vision*. Springer-Verlag, 2001.

[52] R. Benosman, T. Manière, and J. Devars, "Multidirectional stereovision sensor, calibration and scenes reconstruction," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 161–165, 1996.

[53] H. Beyer, "Accurate calibration of CCD cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pp. 96–101, 1992.

[54] S. Bogner, "An introduction to panospheric imaging," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vancouver, Canada*, pp. 3099–3106, 1995.

[55] J. Boland, "Cameras and sensing systems," in *Manual of Photogrammetry, Fifth Edition*, (C. McGlone, ed.), ch. 8, pp. 581–676, Falls Church, Virginia, USA: American Society of Photogrammetry and Remote Sensing, 2004.

[56] R. Bolles, H. Baker, and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7–55, 1987.

[57] R. Boutteau, X. Savatier, J.-Y. Ertaud, and B. Mazari, "An omnidirectional stereoscopic system for mobile robot navigation," in *Proceedings of the IEEE International Workshop on Robotic and Sensors Environment, Ottawa, Canada*, pp. 138–143, 2008.

[58] A. Bouwers and J. van der Sande, "A new camera of extremely high luminosity and resolution based on the concentric optical mirror system," in *Proceedings of the VII ISPRS-Congress, Washington, USA*, pp. 246/1–246/3, 1952.

[59] P. Brand, R. Mohr, and P. Bobet, "Distorsions optiques: Correction dans un modele projectif," Technical Report 1933, LIFIA–IMAG–INRIA Rhône-Alpes, Grenoble, France, June 1993.

[60] P. Brand, R. Mohr, and P. Bobet, "Distorsion optique: Correction dans un modèle projectif," in *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pp. 87–98, Paris, January 1994.

[61] E. Brassart, L. Delahoche, C. Cauchois, C. Drocourt, C. Pegard, and M. Mouaddib, "Experimental results got with the omnidirectional vision sensor: SYCLOP," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, pp. 145–152, 2000.

[62] C. Bräuer-Burchardt and K. Voss, "A new algorithm to correct fish-eye and strong wide-angle-lens-distortion from single images," in *Proceedings of the IEEE International Conference on Image Processing, Thessaloniki, Greece*, pp. 225–228, 2001.

[63] D. Brown, "Decentering distortion of lenses," *Photogrammetric Engineering*, vol. 32, pp. 444–462, May 1966.

[64] D. Brown, "Close-range camera calibration," *Photogrammetric Engineering*, vol. 37, no. 8, pp. 855–866, 1971.

[65] L. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325–376, December 1992.

[66] A. Bruckstein and T. Richardson, "Omniview cameras with curved surface mirrors," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, pp. 79–84, 2000.

[67] T. Buchanan, "The twisted cubic and camera calibration," *Computer Vision, Graphics and Image Processing*, vol. 42, pp. 130–132, April 1988.

[68] R. Bunschoten and B. Kröse, "3D scene reconstruction from cylindrical panoramic images," *Robotics and Autonomous Systems*, vol. 41, no. 2–3, pp. 111–118, 2002.

[69] R. Bunschoten and B. Kröse, "Robust scene reconstruction from an omnidirectional vision system," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 2, pp. 351–357, 2002.

[70] A. Burner, "Zoom lens calibration for wind tunnel measurements," in *Proceedings of the SPIE Conference on Videometrics IV, Philadelphia, Pennsylvania, USA*, (S. El-Hakim, ed.), pp. 19–33, SPIE — Society of Photo-Optical Instrumentation Engineers, October 1995.

[71] P. Burns, "The history of the discovery of cinematography," http://www.precinemahistory.net.

[72] M. Byröd, Z. Kúkelová, K. Josephson, T. Pajdla, and K. Åström, "Fast and robust numerical solutions to minimal problems for cameras with radial distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA*, 2008.

[73] E. Cabral, J. de Souza Junior, and M. Hunold, "Omnidirectional stereo vision with a hyperbolic double lobed mirror," in *Proceedings of the 15th International Conference on Pattern Recognition, Cambridge, UK*, pp. 1–4, 2004.

[74] C. Cafforio and F. Rocca, "Precise stereopsis with a single video camera," in *Proceedings of the 3rd European Signal Processing Conference (EUSIPCO): Theories and Application, The Hague, Netherlands*, pp. 641–644, 1986.

[75] V. Caglioti and S. Gasparini, "Localization of straight lines from single images using off-axis catadioptric cameras," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[76] V. Caglioti and S. Gasparini, "On the localization of straight lines in 3D space from single 2D images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 1129–1134, 2005.

[77] V. Caglioti and S. Gasparini, ""How many planar viewing surfaces are there in noncentral catadioptric cameras?" Towards singe-image localization of space lines," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 1266–1273, 2006.

[78] V. Caglioti, P. Taddei, G. Boracchi, S. Gasparini, and A. Giusti, "Single-image calibration of off-axis catadioptric cameras using lines," in *Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Rio de Janeiro, Brazil*, 2007.

[79] S. Cagnoni, M. Mordonini, L. Mussi, and G. Adorni, "Hybrid stereo sensor with omnidirectional vision capabilities: Overview and calibration procedures," in *Proceedings of the 14th International Conference on Image Analysis and Processing, Modena, Italy*, pp. 99–104, 2007.

[80] Z. Cao, S. Oh, and E. Hall, "Dynamic omnidirectional vision for mobile robots," *Journal of Robotic Systems*, vol. 3, no. 1, pp. 5–17, 1986.

[81] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *International Journal of Computer Vision*, vol. 4, pp. 127–140, 1990.

[82] C. Cauchois, E. Brassart, L. Delahoche, and A. Clerentin, "3D localization with conical vision," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[83] C. Cauchois, E. Brassart, C. Drocourt, and P. Vasseur, "Calibration of the omnidirectional vision sensor: SYCLOP," in *Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, Michigan, USA*, pp. 1287–1292, May 1999.

[84] J. Chahl and M. Srinivasan, "Reflective surfaces for panoramic imaging," *Applied Optics*, vol. 36, no. 31, pp. 8275–8285, 1997.

[85] G. Champleboux, S. Lavallée, P. Sautot, and P. Cinquin, "Accurate calibration of cameras and range imaging sensors: The NPBS method," in *Proceedings of the IEEE International Conference on Robotics and Automation, Nice, France*, pp. 1552–1558, May 1992.

[86] P. Chang and M. Hébert, "Omni-directional structure from motion," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, pp. 127–133, 2000.

[87] A. Charriou and S. Valette, "Appareil photographique multiple pour études sur la prise de vues aériennes," in *Proceedings of the V ISPRS-Congress, Rome, Italy*, pp. 154–157, 1938.

[88] C.-S. Chen and W.-Y. Chang, "On pose recovery for generalized visual sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 848–861, July 2004.

[89] N.-Y. Chen, "Visually estimating workpiece pose in a robot hand using the feature points method," PhD thesis, University of Rhode Island, Kingston, 1979.

[90] N.-Y. Chen, J. Birk, and R. Kelley, "Estimating workpiece pose using the feature points method," *IEEE Transactions on Automatic Control*, vol. 25, pp. 1027–1041, December 1980.

[91] Q. Chen, H. Wu, and T. Wada, "Camera calibration with two arbitrary coplanar circles," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, pp. 521–532, 2004.

[92] S. Chen, "Quicktime VR — an image-based approach to virtual environment navigation," in *Proceedings of SIGGRAPH, Los Angeles, USA*, pp. 29–38, 1995.

[93] Y. Chen and H. Ip, "Single view metrology of wide-angle lens images," *The Visual Computer*, vol. 22, no. 7, pp. 445–455, 2006.

[94] J. Clark and G. Follin, "A simple "equal area" calibration for fisheye photography," *Agricultural and Forest Meteorology*, vol. 44, pp. 19–25, 1988.

[95] T. Clarke and J. Fryer, "The development of camera calibration methods and models," *Photogrammetric Record*, vol. 91, no. 16, pp. 51–66, 1998.

[96] T. Clarke, X. Wang, and J. Fryer, "The principal point and CCD cameras," *Photogrammetric Record*, vol. 92, no. 16, pp. 293–312, 1998.

[97] D. Claus and A. Fitzgibbon, "A plumbline constraint for the rational function lens distortion model," in *Proceedings of the 16th British Machine Vision Conference, Oxford, England*, pp. 99–108, 2005.

[98] D. Claus and A. Fitzgibbon, "A rational function lens distortion model for general cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 213–219, 2005.

[99] T. Conroy and J. Moore, "Resolution invariant surfaces for panoramic vision systems," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece*, pp. 392–397, 1999.

[100] K. Cornelis, M. Pollefeys, and L. van Gool, "Lens distortion recovery for accurate sequential structure and motion recovery," in *Proceedings of*

the 7th European Conference on Computer Vision, Copenhagen, Denmark, pp. 186–200, 2002.

[101] N. Cornille, "Accurate 3D shape and displacement measurement using a scanning electron microscope," PhD thesis, University of South Carolina and Institut National des Sciences Appliquées de Toulouse, June 2005.

[102] J. Crowley, P. Bobet, and C. Schmid, "Maintaining stereo calibration by tracking image points," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 483–488, June 1993.

[103] K. Daniilidis, "The page of omnidirectional vision," http://www.cis.upenn.edu/∼kostas/omni.html.

[104] N. Daucher, M. Dhome, and J.-T. Lapresté, "Camera calibration from sphere images," in *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pp. 449–454, 1994.

[105] A. Davidhazy, "Camera for conical peripheral and panoramic photography," SIGGRAPH Course, 2007.

[106] C. Davis and T.-H. Ho, "Using geometrical constraints for fisheye camera calibration," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[107] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[108] T. Debaecker, R. Benosman, and S. Ieng, "Cone-pixels camera models," in *Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France*, 2008.

[109] C. Delherm, J.-M. Lavest, M. Dhome, and J.-T. Lapresté, "Dense reconstruction by zooming," in *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, (B. Buxton and R. Cipolla, eds.), pp. 427–438, Springer-Verlag, 1996.

[110] X.-M. Deng, F.-C. Wu, and Y.-H. Wu, "An easy calibration method for central catadioptric cameras," *Acta Automatica Sinica*, vol. 33, no. 8, pp. 801–808, 2007.

[111] S. Derrien and K. Konolige, "Approximating a single viewpoint in panoramic imaging devices," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3931–3938, 2000.

[112] R. Descartes, *Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences.* Ian Maire, Leyden, 1637.

[113] F. Devernay and O. Faugeras, "Straight lines have to be straight," *Machine Vision and Applications*, vol. 13, pp. 14–24, August 2001.

[114] M. Dhome, J. Lapreste, G. Rives, and M. Richetin, "Spatial localization of modelled objects of revolution in monocular perspective vision," in *Proceedings of the 1st European Conference on Computer Vision, Antibes, France*, pp. 475–488, 1990.

[115] H. Dietz, "Fisheye digital imaging for under twenty dollars," Technical Report, University of Kentucky, April 2006.

[116] Y. Ding and J. Yu, "Multiperspective distortion correction using collineations," in *Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan*, pp. 95–105, 2007.

[117] E. Doležal, "Photogrammetrische Lösung des Wolkenproblems aus einem Standpunkt unter Verwendung der Reflexe," *Sitzungsberichte Kaiserliche Akademie der Wissenschaften, mathematisch-naturwissenschaftliche Klasse, Abteilung IIa*, vol. 111, pp. 788–813, 1902.

[118] F. Dornaika and J. Elder, "Image registration for foveated omnidirectional sensing," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pp. 606–620, 2002.

[119] P. Doubek and T. Svoboda, "Reliable 3D reconstruction from a few catadioptric images," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, 2002.

[120] J. Draréni, P. Sturm, and S. Roy, "Plane-based calibration for linear cameras," in *Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France*, October 2008.

[121] D. Drucker and P. Locke, "A natural classiffication of curves and surfaces with reflection properties," *Mathematics Magazine*, vol. 69, no. 4, pp. 249–256, 1996.

[122] F. Du and M. Brady, "Self-calibration of the intrinsic parameters of cameras for active vision systems," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 477–482, IEEE Computer Society Press, 1993.

[123] A. Dunne, J. Mallon, and P. Whelan, "A comparison of new generic camera calibration with the standard parametric approach," in *Proceedings of IAPR Conference on Machine Vision Applications, Tokyo, Japan*, pp. 114–117, 2007.

[124] A. Dunne, J. Mallon, and P. Whelan, "Efficient generic calibration method for general cameras with single centre of projection," in *Proceedings of the 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil*, 2007.

[125] A. Dunne, J. Mallon, and P. Whelan, "Efficient generic calibration method for general cameras with single centre of projection," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 220–233, 2010.

[126] T. Echigo, "A camera calibration technique using three sets of parallel lines," *Machine Vision and Applications*, vol. 3, no. 3, pp. 159–167, 1990.

[127] M. El-Melegy and A. Farag, "Nonmetric lens distortion calibration: Closed-form solutions, robust estimation and model selection," in *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France*, pp. 554–559, 2003.

[128] R. Enciso, T. Viéville, and A. Zisserman, "An affine solution to euclidean calibration for a zoom lens," in *Proceedings of the ALCATECH Workshop, Denmark*, pp. 21–27, July 1996.

[129] C. Engels, H. Stewénius, and D. Nistér, "Bundle adjustment rules," in *Proceedings of ISPRS Symposium on Photogrammetric Computer Vision, Bonn, Germany*, 2006.

[130] F. Espuny, "A closed-form solution for the generic self-calibration of central cameras from two rotational flows," in *Proceedings of the International Conference on Computer Vision Theory and Applications, Barcelona, Spain*, pp. 26–31, 2007.

[131] F. Espuny and J. Burgos Gil, "Generic self-calibration of central cameras from two "real" rotational flows," in *Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France*, 2008.

[132] J. Fabrizio and J. Devars, "An analytical solution to the perspective-n-point problem for common planar camera and for catadioptric sensor," *International Journal of Image and Graphics*, vol. 8, pp. 135–155, January 2008.

[133] J. Fabrizio, J. Tarel, and R. Benosman, "Calibration of panoramic catadioptric sensors made easier," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pp. 45–52, 2002.

[134] W. Faig, "Calibration of close-range photogrammetric systems: Mathematical formulation," *Photogrammetric Engineering & Remote Sensing*, vol. 41, pp. 1479–1486, December 1975.

[135] H. Farid and A. Popescu, "Blind removal of image non-linearities," in *Proceedings of the 8th IEEE International Conference on Computer Vision, Vancouver, Canada*, pp. 76–81, 2001.

[136] O. Fassig, "A revolving cloud camera," *Monthly Weather Review*, vol. 43, no. 6, pp. 274–275, 1915.

[137] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The Geometry of Multiple Images*. MIT Press, March 2001.

[138] O. Faugeras and B. Mourrain, "On the geometry and algebra of the point and line correspondences between n images," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pp. 951–956, June 1995.

[139] D. Feldman, T. Pajdla, and D. Weinshall, "On the epipolar geometry of the crossed-slits projection," in *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France*, pp. 988–995, 2003.

[140] R. Fergus, A. Torralba, and W. Freeman, "Random lens imaging," MIT-CSAIL-TR-2006-058, Massachusetts Institute of Technology, September 2006.

[141] C. Fermüller and Y. Aloimonos, "Geometry of eye design: Biology and technology," in *Proceedings of the 10th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany*, pp. 22–38, Springer-Verlag, 2001.

[142] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol. 1, Mainly Mechanics, Radiation, and Heat*. Addison-Wesley, 1963.

[143] M. Fiala and A. Basu, "Feature extraction and calibration for stereo reconstruction using non-svp optics in a panoramic stereo-vision sensor," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, 2002.

[144] S. Finsterwalder, "Die geometrischen Grundlagen der Photogrammetrie," *Jahresbericht Deutscher Mathematik*, vol. 6, pp. 1–44, 1899.

[145] A. Fitzgibbon, "Simultaneous linear estimation of multiple view geometry and lens distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 125–132, 2001.

[146] M. Fleck, "Perspective projection: The wrong imaging model," Technical Report TR 95–01, Department of Computer Science, University of Iowa, Iowa City, IA 52242, USA, 1995.

[147] W. Förstner, B. Wrobel, F. Paderes, R. Craig, C. Fraser, and J. Dolloff, "Analytical photogrammetric operations," in *Manual of Photogrammetry, Fifth Edition*, (C. McGlone, ed.), ch. 11, pp. 763–948, Falls Church, Virginia, USA: American Society of Photogrammetry and Remote Sensing, 2004.

[148] O. Frank, R. Katz, C. Tisse, and H. Durrant-Whyte, "Camera calibration for miniature, low-cost, wide-angle imaging system," in *Proceedings of the 18th British Machine Vision Conference, Warwick, England*, 2007.

[149] C. Fraser, "Digital camera self-calibration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, no. 4, pp. 149–159, 1997.

[150] C. Fraser and S. Al-Ajlouni, "Zoom-dependent calibration for consumer grade-cameras," *Photogrammetric Engineering & Remote Sensing*, vol. 72, pp. 1017–1026, September 2006.

[151] L. Fritz and H. Schmid, "Stellar calibration of the orbigon lens," *Photogrammetric Engineering*, vol. 40, no. 1, pp. 101–115, 1974.

[152] J. Fryer and D. Brown, "Lens distortion for close-range photogrammetry," *Photogrammetric Engineering & Remote Sensing*, vol. 52, pp. 51–58, January 1986.

[153] S. Gächter, T. Pajdla, and B. Mičušík, "Mirror design for an omnidirectional camera with a space variant imager," in *Proceedings of the Workshop on Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing, Budapest, Hungary*, 2001.

[154] C. Gao and N. Ahuja, "Single camera stereo using planar parallel plate," in *Proceedings of the 15th International Conference on Pattern Recognition, Cambridge, UK*, pp. 108–111, 2004.

[155] C. Gao, H. Hua, and N. Ahuja, "A hemispherical imaging camera," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 168–178, 2010.

[156] J. Gaspar, "Omnidirectional vision for mobile robot navigation," PhD thesis, Universidade Técnica de Lisboa, Portgual, December 2002.

[157] J. Gaspar, C. Deccó, J. Okamoto Jr, and J. Santos-Victor, "Constant resolution omnidirectional cameras," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pp. 27–34, 2002.

[158] S. Gasparini and P. Sturm, "Multi-view matching tensors from lines for general camera models," in *Tensors in Image Processing and Computer Vision*, (S. Aja-Fernández, R. de Luis García, D. Tao, and X. Li, eds.), Springer-Verlag, 2009.

[159] S. Gasparini, P. Sturm, and J. Barreto, "Plane-based calibration of central catadioptric cameras," in *Proceedings of the 12th IEEE International Conference on Computer Vision, Kyoto, Japan*, 2009.

[160] M. Gasser, "Mehrfachkammer für Aufnahmen aus Luftfahrzeugen," Patent No. 469,413, Reichspatentamt, Germany, 1926.

[161] D. Gennery, "Generalized camera calibration including fish-eye lenses," *International Journal of Computer Vision*, vol. 68, pp. 239–266, July 2006.

[162] T. Georgiev, C. Intwala, and D. Babacan, "Light-field capture by multiplexing in the frequency domain," Technical Report, Adobe Systems Incorporated, April 2007.

[163] C. Geyer and K. Daniilidis, "Catadioptric camera calibration," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece*, pp. 398–404, 1999.

[164] C. Geyer and K. Daniilidis, "A unifying theory of central panoramic systems and practical applications," in *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, (D. Vernon, ed.), pp. 445–461, Springer-Verlag, June 2000.

[165] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223–243, 2001.

[166] C. Geyer and K. Daniilidis, "Structure and motion from uncalibrated catadioptric views," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 279–286, 2001.

[167] C. Geyer and K. Daniilidis, "Paracatadioptric camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 687–695, May 2002.

[168] C. Geyer and K. Daniilidis, "Properties of the catadioptric fundamental matrix," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, (A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, eds.), pp. 140–154, 2002.

[169] C. Geyer and K. Daniilidis, "Conformal rectification of an omnidirectional stereo pair," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[170] C. Geyer and K. Daniilidis, "Mirrors in motion: Epipolar geometry and motion estimation," in *Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France*, pp. 766–773, 2003.

[171] C. Geyer, M. Meingast, and S. Sastry, "Geometric models of rolling-shutter cameras," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[172] C. Geyer and H. Stewénius, "A nine-point algorithm for estimating paracatadioptric fundamental matrices," in *Proceedings of the 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil*, pp. 1–8, 2007.

[173] J. Gluckman and S. Nayar, "Ego-motion and omnidirectional cameras," in *Proceedings of the 6th IEEE International Conference on Computer Vision, Bombay, India*, pp. 999–1005, January 1998.

[174] J. Gluckman and S. Nayar, "Planar catadioptric stereo: geometry and calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 22–28, 1999.

[175] J. Gluckman and S. Nayar, "Catadioptric stereo using planar mirrors," *International Journal of Computer Vision*, vol. 44, pp. 65–79, August 2001.

[176] J. Gluckman and S. Nayar, "Rectifying transformations that minimize resampling effects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 111–117, 2001.

[177] J. Gluckman and S. Nayar, "Rectified catadioptric stereo sensors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 224–236, February 2002.

[178] S. Godber, R. Petty, M. Robinson, and J. Evans, "Panoramic line-scan imaging system for teleoperator control," in *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems*, pp. 247–257, 1994.

[179] N. Gonçalvez and H. Araújo, "Projection model, 3D reconstruction and rigid motion estimation from non-central catadioptric images," in *Proceedings of the Second International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill, USA*, pp. 325–332, 2004.

[180] N. Gonçalvez and H. Araújo, "Estimating parameters of noncentral catadioptric systems using bundle adjustment," *Computer Vision and Image Understanding*, vol. 113, pp. 11–28, January 2009.

[181] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proceedings of SIGGRAPH, New Orleans, LA*, pp. 43–54, 1996.

[182] A. Goshtasby, "Correction of image deformation from lens distortion using Bezier patches," *Computer Vision, Graphics and Image Processing*, vol. 47, pp. 385–394, 1989.

[183] A. Goshtasby and W. Gruver, "Design of a single-lens stereo camera system," *Pattern Recognition*, vol. 26, no. 6, pp. 923–938, 1993.

[184] S. Gourichon, J. Meyer, S. Ieng, L. Smadja, and R. Benosman, "Estimating ego-motion using a panoramic sensor: Comparison between a bio-inspired and a camera-calibrated method," in *Proceedings of the AISB Symposium on Biologically Inspired Vision, Theory and Application*, pp. 91–101, 2003.

[185] G. Gracie, "Analytical photogrammetry applied to single terrestrial photograph mensuration," in *Proceedings of the XIth International Congress of Photogrammetry, Lausanne, Switzerland*, July 1968.

[186] W. Green, P. Jepsen, J. Kreznar, R. Ruiz, A. Schwartz, and J. Seidman, "Removal of instrument signature from mariner 9 television images of mars," *Applied Optics*, vol. 14, no. 1, pp. 105–114, 1975.

[187] N. Greene, "Environment mapping and other applications of world projections," *IEEE Transactions on Computer Graphics and Applications*, vol. 6, no. 11, pp. 21–29, 1986.

[188] P. Greguss, "The tube peeper: A new concept in endoscopy," *Optics & Laser Technology*, vol. 17, no. 1, pp. 41–45, 1985.

[189] K. Gremban, C. Thorpe, and T. Kanade, "Geometric camera calibration using systems of linear equations," in *Proceedings of the IEEE International Conference on Robotics and Automation, Philadelphia, Pennsylvania, USA*, pp. 562–567, IEEE Computer Society Press, 1988.

[190] W. Grosky and L. Tamburino, "A unified approach to the linear camera calibration problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 663–671, July 1990.

[191] M. Grossberg and S. Nayar, "The raxel imaging model and ray-based calibration," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 119–137, 2005.

[192] E. Grossmann, J. Gaspar, and F. Orabona, "Discrete camera calibration from pixel streams," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 198–209, 2010.

[193] E. Grossmann, E.-J. Lee, P. Hislop, D. Nistér, and H. Stewénius, "Are two rotational flows sufficient to calibrate a smooth non-parametric sensor?" in

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, pp. 1222–1229, 2006.

[194] E. Grossmann, F. Orabona, and J. Gaspar, "Discrete camera calibration from the information distance between pixel streams," in *Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Rio de Janeiro, Brazil*, 2007.

[195] A. Gruen and T. Huang, eds., *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag, 2001.

[196] X. Gu, S. Gortler, and M. Cohen, "Polyhedral geometry and the two-plane parameterization," in *Proceedings of the Eurographics Workshop on Rendering Techniques, St. Etienne, France*, pp. 1–12, 1997.

[197] R. Gupta and R. Hartley, "Linear pushbroom cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 963–975, September 1997.

[198] R. Gupta and R. Hartley, "Camera estimation for orbiting pushbroom imaging systems," unpublished, 2007.

[199] P. Gurdjos, P. Sturm, and Y. Wu, "Euclidean structure from $n >= 2$ parallel circles: Theory and algorithms," in *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, (H. Bischof and A. Leonardis, eds.), pp. 238–252, May 2006.

[200] P. Hall, J. Collomosse, Y. Song, P. Shen, and C. Li, "RTcams: A new perspective on non-photorealistic rendering from photographs," *IEEE Transactions on Visualization and Graphics*, vol. 13, no. 5, pp. 966–979, 2007.

[201] B. Hallert, "A new method for the determination of the distortion and the inner orientation of cameras and projectors," *Photogrammetria*, vol. 11, pp. 107–115, 1954–1955.

[202] B. Hallert, "The method of least squares applied to multicollimator camera calibration," *Photogrammetric Engineering*, vol. 29, no. 5, pp. 836–840, 1963.

[203] J. Han and K. Perlin, "Measuring bidirectional texture reflectance with a kaleidoscope," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 741–748, 2003.

[204] H. Haneishi, Y. Yagihashi, and Y. Miyake, "A new method for distortion correction of electronic endoscope images," *IEEE Transactions on Medical Imaging*, vol. 14, pp. 548–555, September 1995.

[205] R. Haralick, C. Lee, K. Ottenberg, and M. Nolle, "Review and analysis of solutions of the three point perspective pose estimation problem," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 331–356, 1994.

[206] R. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, (G. Sandini, ed.), pp. 579–587, Springer-Verlag, 1992.

[207] R. Hartley, "An algorithm for self calibration from several views," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pp. 908–912, 1994.

[208] R. Hartley, "In defence of the 8-point algorithm," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pp. 1064–1070, June 1995.

[209] R. Hartley and S. Kang, "Parameter-free radial distortion correction with center of distortion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1309–1321, 2007.

[210] R. Hartley and T. Saxena, "The cubic rational polynomial camera model," in *Proceedings of the DARPA Image Understanding Workshop, New Orleans, Louisiana, USA*, pp. 649–653, 1997.

[211] R. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[212] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.

[213] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd Edition, March 2004.

[214] J. Havlena, A. Torii, J. Knopp, and T. Pajdla, "Randomized structure from motion based on atomic 3D models from camera triplets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA*, pp. 2874–2881, 2009.

[215] J. Heikkilä, "Geometric camera calibration using circular control points," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1066–1077, 2000.

[216] J. Heikkilä and O. Silvén, "Calibration procedure for short focal length off-the-shelf CCD cameras," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 166–170, IEEE Computer Society Press, August 1996.

[217] J. Heller and T. Pajdla, "Stereographic rectification of omnidirectional stereo pairs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA*, pp. 1414–1421, 2009.

[218] H.-J. Hellmeier, "Fisheye-Objektive in der Nahbereichsphotogrammetrie — Theoretische und praktische Untersuchungen," PhD thesis, Technische Universität Braunschweig, Germany, 1983.

[219] T. Herbert, "Calibration of fisheye lenses by inversion of area projections," *Applied Optics*, vol. 25, no. 12, pp. 1875–1876, 1986.

[220] R. Hicks, "The page of catadioptric sensor design," http://www.math.drexel.edu/ ahicks/design/.

[221] R. Hicks, "Designing a mirror to realize a given projection," *Journal of the Optical Society of America A*, vol. 22, no. 2, pp. 323–330, 2005.

[222] R. Hicks and R. Bajcsy, "Catadioptric sensors that approximate wide-angle perspective projections," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pp. 545–551, 2000.

[223] R. Hicks and R. Bajcsy, "Reflective surfaces as computational sensors," *Image and Vision Computing*, vol. 19, no. 11, pp. 773–777, 2001.

[224] R. Hicks, V. Nasis, and T. Kurzweg, "Programmable imaging with two-axis micromirrors," *Optics Letters*, vol. 32, no. 9, pp. 1066–1068, 2007.

[225] R. Hicks and R. Perline, "Equi-areal catadioptric sensors," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pp. 13–18, 2002.

[226] R. Hill, "A lens for whole sky photographs," *Quarterly Journal of the Royal Meteorological Society*, vol. 50, no. 211, pp. 227–235, 1924.

[227] historiccamera.com, "Illustrated history of photography," http://www.historiccamera.com/history1/photo_history300.html.

[228] S. Hiura, A. Mohan, and R. Raskar, "Krill-eye: Superposition compound eye for wide-angle imaging via GRIN lenses," in *Proceedings of the 9th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Kyoto, Japan*, 2009.

[229] O. Holmes, "The stereoscope and the stereograph," *The Atlantic Monthly*, vol. 3, pp. 738–749, June 1859.

[230] R. Holt and A. Netravali, "Camera calibration problem: Some new results," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 54, pp. 368–383, November 1991.

[231] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. Riseman, "Image-based homing," in *Proceedings of the IEEE International Conference on Robotics and Automation, Sacramento, California, USA*, pp. 620–625, April 1991.

[232] R. Horaud, F. Dornaika, B. Lamiroy, and S. Christy, "Object pose: The link between weak perspective, paraperspective and full perspective," *International Journal of Computer Vision*, vol. 22, pp. 173–189, March 1997.

[233] B. Horn, H. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America A*, vol. 5, pp. 1127–1135, July 1988.

[234] H. Hua and N. Ahuja, "A high-resolution panoramic camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 960–967, 2001.

[235] H. Hua, N. Ahuja, and C. Gao, "Design analysis of a high-resolution panoramic camera using conventional imagers and a mirror pyramid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 356–361, 2007.

[236] F. Huang, "Epipolar geometry in concentric panoramas," Technical Report CTU-CMP-2000-07, Center for Machine Perception, Czech Technical University, Prague, 2000.

[237] F. Huang, R. Klette, and Y.-H. Xie, "Sensor pose estimation from multi-center cylindrical panoramas," in *Proceedings of the Third Pacific Rim Symposium on Advances in Image and Video Technology, Tokyo, Japan*, pp. 48–59, 2008.

[238] F. Huang, S. Wei, and R. Klette, "Epipolar geometry in polycentric panoramas," in *Proceedings of the 10th International Workshop on Theoretical Foundations of Computer Vision, Dagstuhl Castle, Germany*, (R. Klette, T. Huang, and G. Gimel'farb, eds.), pp. 39–50, Springer-Verlag, 2000.

[239] F. Huang, S.-K. Wei, and R. Klette, "Comparative studies of line-based panoramic camera calibration," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[240] C. Hughes, P. Denny, M. Glavin, and E. Jones, "Equidistant fish-eye calibration and rectification by vanishing point extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2289–2296, 2010.

[241] C. Hughes, R. McFeely, P. Denny, M. Glavin, and E. Jones, "Equidistant ($f\theta$) fish-eye perspective with application in distortion centre estimation," *Image and Vision Computing*, vol. 28, no. 3, pp. 538–551, 2010.

[242] Y. Hwang, J. Lee, and H. Hong, "Omnidirectional camera calibration and 3D reconstruction by contour matching," in *Proceedings of the Second International Symposium on Visual Computing, Lake Tahoe, USA*, pp. 881–890, 2006.

[243] N. Ichimura and S. K. Nayar, "A framework for 3D pushbroom imaging," Technical Report CUCS-002-03, Department of Computer Science, Columbia University, 2003.

[244] M. Inaba, T. Hara, and H. Inoue, "A stereo viewer based on a single camera with view-control mechanisms," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan*, pp. 1857–1864, 1993.

[245] A. Inoue, K. Yamamoto, N. Mizoue, and Y. Kawahara, "Calibrating view angle and lens distortion of the Nikon fish-eye converter FC-E8," *Journal for Forestry Research*, vol. 9, pp. 177–181, 2004.

[246] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo for making global map," in *Proceedings of the 3rd IEEE International Conference on Computer Vision, Osaka, Japan*, pp. 540–547, 1990.

[247] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 257–262, 1992.

[248] F. Ives, "Parallax stereogram and process of making same, U.S. Patent 725,567," 1903.

[249] H. Ives, "A camera for making parallax panoramagrams," *Journal of the Optical Society of America*, vol. 17, pp. 435–437, 1928.

[250] U. Iwerks, "Panoramic motion picture camera arrangement, U.S. Patent 3,118, 340," 1964.

[251] A. Izaguirre, P. Pu, and J. Summers, "A new development in camera calibration — calibrating a pair of mobile cameras," in *Proceedings of the IEEE International Conference on Robotics and Automation, Saint Louis, Michigan, USA*, pp. 74–79, 1985.

[252] M. Jackowski, A. Goshtasby, S. Bines, D. Roseman, and C. Yu, "Correcting the geometry and color of digital images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 1152–1158, October 1997.

[253] B. Jähne, *Digitale Bildverarbeitung*. Springer-Verlag, 1st Edition, 1989.

[254] B. Jähne, *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer-Verlag, 1st Edition, 1991.

[255] G. Jang, S. Kim, and I. Kweon, "Single camera catadioptric stereo system," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[256] S. Jeng and W. Tsai, "Analytic image unwarping by a systematic calibration method for omni-directional cameras with hyperbolic-shaped mirrors," *Image and Vision Computing*, vol. 26, pp. 690–701, May 2008.

[257] G. Jiang, H.-T. Tsui, L. Quan, and A. Zisserman, "Single axis geometry by fitting conics," in *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pp. 537–550, 2002.

[258] A. Jones, P. Debevec, M. Bolas, and I. McDowall, "Concave surround optics for rapid multi-view imaging," in *Proceedings of the 25th Army Science Conference, Orlando, USA*, 2006.

[259] K. Josephson and M. Byröd, "Pose estimation with radial distortion and unknown focal length," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA*, 2009.

[260] F. Kahl, S. Agarwal, M. Chandraker, D. Kriegman, and S. Belongie, "Practical global optimization for multiview geometry," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 271–284, 2008.

[261] S. Kaneko and T. Honda, "Calculation of polyhedral objects using direct and mirror images," *Journal of the Japan Society for Precision Engineering*, vol. 52, no. 1, pp. 149–155, 1986.

[262] S. Kang, "Catadioptric self-calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pp. 201–207, 2000.

[263] S. Kang, "Radial distortion snakes," *IEICE Transactions on Information and Systems*, vol. E84-D, no. 12, pp. 1603–1611, 2001.

[264] S. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 167–183, 1997.

[265] F. Kangni and R. Laganière, "Epipolar geometry for the rectification of cubic panoramas," in *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision, Québec City, Canada*, 2006.

[266] J. Kannala and S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1335–1340, 2006.

[267] J. Kannala, S. Brandt, and J. Heikkilä, "Self-calibration of central cameras by minimizing angular error," in *Proceedings of the International Conference on Computer Vision Theory and Applications, Funchal, Portugal*, 2008.

[268] U.-P. Käppeler, M. Höferlin, and P. Levi, "3D object localization via stereo vision using an omnidirectional and a perspective camera," in *Proceedings of the 2nd Workshop on Omnidirectional Robot Vision, Anchorage, Alaska*, pp. 7–12, 2010.

[269] R. Karren, "Camera calibration by the multicollimator method," *Photogrammetric Engineering*, vol. 34, no. 7, pp. 706–719, 1968.

[270] K. Kato, T. Nakanishi, A. Shio, and K. Ishii, "Structure from image sequences captured through a monocular extra-wide angle lens," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA*, pp. 919–924, 1994.

[271] T. Kawanishi, K. Yamazawa, H. Iwasa, H. Takemura, and N. Yokoya, "Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors," in *Proceedings of the 14th International Conference on Pattern Recognition, Brisbane, Australia*, pp. 485–489, 1998.

[272] M. Kedzierski and A. Fryskowska, "Precise method of fisheye lens calibration," in *Proceedings of the ISPRS-Congress, Beijing, China*, pp. 765–768, 2008.

[273] S. Khan, F. Rafi, and M. Shah, "Where was the picture taken: Image localization in route panoramas using epipolar geometry," in *Proceedings of the IEEE International Conference on Multimedia and Expo, Toronto, Canada*, pp. 249–252, 2006.

[274] E. Kilpelä, "Compensation of systematic errors of image and model coordinates," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXIII, no. B9, pp. 407–427, 1980.

[275] E. Kilpelä, "Compensation of systematic of image and model coordinates," *Photogrammetria*, vol. 37, no. 1, pp. 15–44, 1981.

[276] E. Kilpelä, J. Heikkilä, and K. Inkilä, "Compensation of systematic errors in bundle adjustment," *Photogrammetria*, vol. 37, no. 1, pp. 1–13, 1981.

[277] J.-H. Kim, H. Li, and R. Hartley, "Motion estimation for nonoverlapping multicamera rigs: Linear algebraic and l_∞ geometric solutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1044–1059, 2010.

[278] J.-S. Kim, M. Hwangbo, and T. Kanade, "Spherical approximation for multiple cameras in motion estimation: Its applicability and advantages," *Computer Vision and Image Understanding*, vol. 114, no. 10, pp. 1068–1083, 2010.

[279] J.-S. Kim and T. Kanade, "Degeneracy of the linear seventeen-point algorithm for generalized essential matrix," *Journal of Mathematical Imaging and Vision*, vol. 37, no. 1, pp. 40–48, 2010.

[280] W. Kim and H. Cho, "Learning-based constitutive parameters estimation in an image sensing system with multiple mirrors," *Pattern Recognition*, vol. 33, no. 7, pp. 1199–1217, 2000.

[281] R. Klette, G. Gimel'farb, S. Wei, F. Huang, K. Scheibe, M. Scheele, A. Börner, and R. Reulke, "On design and applications of cylindrical panoramas," in *Proceedings of the 10th International Conference on Computer Analysis of Images and Patterns, Groningen, The Netherlands*, pp. 1–8, Springer-Verlag, 2003.

[282] Y. Kojima, R. Sagawa, T. Echigo, and Y. Yagi, "Calibration and performance evaluation of omnidirectional sensor with compound spherical mirrors," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[283] O. Kölbl, "Analytische Verzeichnungsdarstellung bei der vollständigen Kalibrierung," *Bildmessung und Luftbildwesen*, vol. 39, no. 4, pp. 169–176, 1971.

[284] K. Kondo, Y. Mukaigawa, T. Suzuki, and Y. Yagi, "Evaluation of HBP mirror system for remote surveillance," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China*, pp. 3454–3461, 2006.

[285] K. Kondo, Y. Mukaigawa, and Y. Yagi, "Free-form mirror design inspired by photometric stereo," in *Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France*, 2008.

[286] K. Kondo, Y. Yagi, and M. Yachida, "Non-isotropic omnidirectional imaging system for an autonomous mobile robot," in *Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain*, pp. 1228–1233, 2005.

[287] K. Kraus, *Photogrammetry: Geometry from Images and Laser Scans*. Walter de Gruyter, Berlin, 2nd Edition, 2007.

[288] A. Krishnan and N. Ahuja, "Panoramic image acquisition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pp. 379–384, 1996.

[289] G. Krishnan and S. Nayar, "Cata-fisheye camera for panoramic imaging," in *Proceedings of the IEEE Workshop on Applications of Computer Vision*, 2008.

[290] G. Krishnan and S. Nayar, "Towards a true spherical camera," in *Proceedings of the SPIE — Human Vision and Electronic Imaging XIV*, 2009.

[291] Z. Kúkelová, M. Byröd, K. Josephson, T. Pajdla, and K. Åström, "Fast and robust numerical solutions to minimal problems for cameras with radial distortion," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 234–244, 2010.

[292] Z. Kúkelová and T. Pajdla, "A minimal solution to the autocalibration of radial distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA*, 2007.

[293] Z. Kúkelová and T. Pajdla, "Two minimal problems for cameras with radial distortion," in *Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Rio de Janeiro, Brazil*, 2007.

[294] J. Kumler and M. Bauer, "Fish-eye lens designs and their relative performance," in *Proceedings of the SPIE Conference on Current Developments in Lens Design and Optical Systems Engineering, San Diego, USA*, pp. 360–369, 2000.

[295] S. Kuthirummal and S. Nayar, "Multiview radial catadioptric imaging for scene capture," *ACM Transactions on Graphics*, vol. 25, no. 3, pp. 916–923, 2006.

[296] S. Kuthirummal and S. Nayar, "Flexible mirror imaging," in *Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Rio de Janeiro, Brazil*, 2007.

[297] G.-I. Kweon, H.-B. Seung, G.-H. Kim, S.-C. Yang, and Y.-H. Lee, "Wide-angle catadioptric lens with a rectilinear projection scheme," *Applied Optics*, vol. 45, no. 34, pp. 8659–8673, 2006.

[298] D. Lanman, D. Crispell, M. Wachs, and G. Taubin, "Spherical catadioptric arrays: Construction, multi-view geometry, and calibration," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission, Chapel Hill, USA*, pp. 81–88, 2006.

[299] D. Lanman, M. Wachs, G. Taubin, and F. Cukierman, "Reconstructing a 3D line from a single catadioptric image," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission, Chapel Hill, USA*, pp. 89–96, 2006.

[300] J.-M. Lavest, C. Delherm, B. Peuchot, and N. Daucher, "Implicit reconstruction by zooming," *Computer Vision and Image Understanding*, vol. 66, no. 3, pp. 301–315, 1997.

[301] J.-M. Lavest, B. Peuchot, C. Delherm, and M. Dhome, "Reconstruction by zooming from implicit calibration," in *Proceedings of the 1st IEEE International Conference on Image Processing*, pp. 1012–1016, 1994.

[302] J.-M. Lavest, M. Viala, and M. Dhome, "Do we really need an accurate calibration pattern to achieve a reliable camera calibration?," in *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, pp. 158–174, May 1998.

[303] D. Lee, I. Kweon, and R. Cipolla, "A biprism-stereo camera system," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 82–87, 1999.

[304] J. Lee, S. You, and U. Neumann, "Large motion estimation for omnidirectional vision," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, 2000.

[305] R. Lenz and R. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology," in *Proceedings of the IEEE International Conference on Robotics and Automation, Raleigh, North Carolina, USA*, pp. 68–75, 1987.

[306] R. Lenz and R. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 713–720, September 1988.

[307] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the SIGGRAPH, New Orleans, LA*, pp. 31–42, 1996.

[308] M. Lhuillier, "Automatic structure and motion using a catadioptric camera," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[309] M. Lhuillier, "Effective and generic structure from motion using angular error," in *Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong*, pp. 67–70, 2006.

[310] M. Lhuillier, "Automatic scene structure and camera motion using a catadioptric system," *Computer Vision and Image Understanding*, vol. 109, pp. 186–203, February 2008.

[311] H. Li and R. Hartley, "A non-iterative method for correcting lens distortion from nine point correspondences," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[312] H. Li and R. Hartley, "Plane-based calibration and auto-calibration of a fisheye camera," in *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, pp. 21–30, 2006.

[313] M. Li and J.-M. Lavest, "Some aspects of zoom lens camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1105–1110, November 1996.

[314] Y. Li, H.-Y. Shum, C.-K. Tang, and R. Szeliski, "Stereo reconstruction from multiperspective panoramas," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 45–62, 2004.

[315] D. Lichti and M. Chapman, "CCD camera calibration using the finite element method," in *Proceedings of the SPIE Conference on Videometrics IV,*

Philadelphia, Pennsylvania, USA, (S. El-Hakim, ed.), SPIE - Society of Photo-Optical Instrumentation Engineers, October 1995.

[316] D. Lichti and M. Chapman, "Constrained FEM self-calibration," *Photogrammetric Engineering & Remote Sensing*, vol. 63, pp. 1111–1119, September 1997.

[317] J. Lim and N. Barnes, "Estimation of the epipole using optical flow at antipodal points," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 245–253, 2010.

[318] J. Lim, N. Barnes, and H. Li, "Estimating relative camera motion from the antipodal-epipolar constraint," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1907–1914, 2010.

[319] S.-S. Lin and R. Bajcsy, "Single-view-point omnidirectional catadioptric cone mirror imager," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 840–845, May 2006.

[320] S.-S. Lin and R. Bajcsy, "Single-viewpoint, catadioptric cone mirror omnidirectional imaging theory and analysis," *Journal of the Optical Society of America A*, vol. 23, no. 12, pp. 2997–3015, 2006.

[321] S.-S. Lin and R. Bajczy, "High resolution catadioptric omni-directional stereo sensor for robot vision," in *Proceedings of the IEEE International Conference on Robotics and Automation, Taipei, Taiwan*, pp. 1694–1699, 2003.

[322] A. Lippman, "Movie-maps: An application of the optical videodisc to computer graphics," *ACM Transactions on Graphics*, vol. 14, no. 3, pp. 32–42, 1980.

[323] G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *Journal de Physique*, vol. 7, pp. 821–825, 1908.

[324] H. Longuet-Higgins, "A computer program for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, September 1981.

[325] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 125–131, 1999.

[326] T. Luhmann, "A historical review on panoramic photogrammetry," in *Proceedings of the ISPRS Workshop on Panorama Photogrammetry, Dresden, Germany*, 2004.

[327] L. Ma, Y. Chen, and K. Moore, "Rational radial distortion models of camera lenses with analytical solution for distortion correction," *International Journal of Information Acquisition*, vol. 1, no. 2, pp. 135–147, 2004.

[328] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry, *An Invitation to 3-D Vision — From Images to Geometric Models*. Springer-Verlag, 2005.

[329] A. Majumder, W. Seales, M. Gopi, and H. Fuchs, "Immersive teleconferencing: A new algorithm to generate seamless panoramic video imagery," in *Proceedings of the seventh ACM international conference on Multimedia, Orlando, USA*, pp. 169–178, 1999.

[330] J. Mallon and P. Whelan, "Precise radial un-distortion of images," in *Proceedings of the 15th International Conference on Pattern Recognition, Cambridge, UK*, pp. 18–21, 2004.

[331] H. Martins, J. Birk, and R. Kelley, "Camera models based on data from two calibration planes," *Computer Graphics and Image Processing*, vol. 17, pp. 173–180, 1981.

[332] T. Mashita, Y. Iwai, and M. Yachida, "Calibration method for misaligned catadioptric camera," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[333] H. Mathieu and F. Devernay, "Système de miroirs pour la stéréoscopie," Technical Report Rapport Technique 0172, INRIA, 1995.

[334] M. Maurette, "Mars rover autonomous navigation," *Autonomous Robots*, vol. 14, pp. 199–208, March 2003.

[335] B. McBride, "A timeline of panoramic cameras," http://www.panoramic photo.com/timeline.htm.

[336] C. McGlone, ed., *Manual of Photogrammetry*. Falls Church, Virginia, USA: American Society of Photogrammetry and Remote Sensing, 5th Edition, 2004.

[337] G. McLean, "Image warping for calibration and removal of lens distortion," in *Proceedings of the IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, Victoria, Canada*, pp. 170–173, 1993.

[338] L. McMillan and G. Bishop, "Plenoptic modeling: An image-based rendering system," in *Proceedings of the SIGGRAPH, Los Angeles, USA*, pp. 39–46, 1995.

[339] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proceedings of the IEEE International Conference on Robotics and Automation, Rome, Italy*, pp. 3945–3950, April 2007.

[340] M. Meingast, C. Geyer, and S. Sastry, "Geometric models of rolling-shutter cameras," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[341] E. Menegatti, "Omnidirectional Vision for Mobile Robots," PhD thesis, Università di Padova, Italy, December 2002.

[342] M. Menem and T. Pajdla, "Constraints on perspective images and circular panoramas," in *Proceedings of the 15th British Machine Vision Conference, Kingston upon Thames, England*, 2004.

[343] H. Mitsumoto, S. Tamura, K. Okazaki, N. Kajimi, and Y. Fukui, "3-D reconstruction using mirror images based on a plane symmetry recovering method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 941–946, September 1992.

[344] B. Mičušik, "Two-view geometry of omnidirectional cameras," PhD thesis, Faculty of Electrical Engineering, Czech Technical University, Prague, June 2004.

[345] B. Mičušik, D. Martinec, and T. Pajdla, "3D metric reconstruction from uncalibrated omnidirectional images," in *Proceedings of the Asian Conference on Computer Vision, Jeju Island, Korea*, 2004.

[346] B. Mičušik and T. Pajdla, "Estimation of omnidirectional camera model from epipolar geometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, 2003.

[347] B. Mičušik and T. Pajdla, "Autocalibration and 3D reconstruction with non-central catadioptric cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA*, pp. 58–65, 2004.

[348] B. Mičušik and T. Pajdla, "Para-catadioptric camera auto-calibration from epipolar geometry," in *Proceedings of the Asian Conference on Computer Vision, Jeju Island, Korea*, 2004.

[349] B. Mičušik and T. Pajdla, "Structure from motion with wide circular field of view cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1–15, 2006.

[350] K. Miyamoto, "Fish eye lens," *Journal of the Optical Society of America*, vol. 54, no. 8, pp. 1060–1061, 1964.

[351] P. Moëssard, *Le cylindrographe, appareil panoramique.* Gauthier-Villars et fils, Paris, 1889.

[352] O. Morel, R. Seulin, and D. Fofi, "Catadioptric camera calibration by polarization imaging," in *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis, Girona, Spain*, pp. 396–403, 2007.

[353] T. Morita, Y. Yasukawa, Y. Inamoto, U. Takashi, and S. Kawakami, "Measurement in three dimensions by motion stereo and spherical mapping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California, USA*, pp. 422–428, 1989.

[354] E. Mouaddib, R. Sagawa, T. Echigo, and Y. Yagi, "Stereovision with a single camera and multiple mirrors," in *Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain*, pp. 800–805, 2005.

[355] E. Mouaddib, R. Sagawa, T. Echigo, and Y. Yagi, "Two or more mirrors for the omnidirectional stereovision?," in *Proceedings of the 2nd IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing, Marrakech, Morocco*, 2006.

[356] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Real-time localization and 3D reconstruction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 363–370, June 2006.

[357] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Generic and real-time structure from motion," in *Proceedings of the 18th British Machine Vision Conference, Warwick, England*, 2007.

[358] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd, "Generic and real-time structure from motion using local bundle adjustment," *Image and Vision Computing*, vol. 27, no. 8, pp. 1178–1193, 2009.

[359] R. Munjy, "Calibrating non-metric cameras using the finite-element method," *Photogrammetric Engineering & Remote Sensing*, vol. 52, pp. 1201–1205, August 1986.

[360] R. Munjy, "Self-calibration using the finite element approach," *Photogrammetric Engineering & Remote Sensing*, vol. 52, pp. 411–418, March 1986.

[361] J. Murphy, "Application of panospheric imaging to a teleoperated lunar rover," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Vancouver, Canada*, pp. 3117–3121, 1995.

[362] D. Murray, "Recovering range using virtual multi-camera stereo," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 285–291, 1995.

[363] H. Nagahara, Y. Yagi, and M. Yachida, "Super wide field of view head mounted display using catadioptrical optics," *Presence*, vol. 15, no. 5, pp. 588–598, 2006.

[364] H. Nagahara, K. Yoshida, and M. Yachida, "An omnidirectional vision sensor with single view and constant resolution," in *Proceedings of the 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil*, 2007.

[365] V. Nalwa, "A true omnidirectional viewer," Technical Report Bell Laboratories Technical Memorandum, BL0115500-960115-01, AT&T Bell Laboratories, 1996.

[366] S. Nayar, "Sphereo: Determining depth using two specular spheres and a single camera," in *Proceedings of the SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision III, Cambridge, USA*, pp. 245–254, November 1988.

[367] S. Nayar, "Catadioptric omnidirectional camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pp. 482–488, 1997.

[368] S. Nayar, "Omnidirectional vision," in *Proceedings of the Eight International Symposium on Robotics Research, Shonan, Japan*, October 1997.

[369] S. Nayar, "Computational cameras: Redefining the image," *Computer*, vol. 39, no. 8, pp. 30–38, 2006.

[370] S. Nayar, V. Branzoi, and T. Boult, "Programmable imaging: Towards a flexible camera," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 7–22, 2006.

[371] S. Nayar and A. Karmarkar, "360 × 360 mosaics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pp. 380–387, 2000.

[372] S. Nayar and V. Peri, "Folded catadioptric cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 217–223, 1999.

[373] R. Nelson and J. Aloimonos, "Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head)," *Biological Cybernetics*, vol. 58, no. 4, pp. 261–273, 1988.

[374] S. Nene and S. Nayar, "Stereo with mirrors," in *Proceedings of the 6th IEEE International Conference on Computer Vision, Bombay, India*, pp. 1087–1094, January 1998.

[375] J. Neumann, C. Fermüller, and Y. Aloimonos, "Polydioptric camera design and 3D motion estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, pp. 294–301, 2003.

[376] Y. Nishimoto and Y. Shirai, "A feature-based stereo model using small disparities," in *Proceedings of the IEEE International Workshop on Industrial Applications of Machine Vision and Machine Intelligence, Tokyo, Japan*, pp. 192–196, 1987.

[377] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 756–770, June 2004.

[378] D. Nistér, "A minimal solution to the generalized 3-point pose problem," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA*, pp. 560–567, 2004.

[379] D. Nistér and H. Stewénius, "A minimal solution to the generalised 3-point pose problem," *Journal of Mathematical Imaging and Vision*, vol. 27, no. 1, pp. 67–79, 2007.

[380] D. Nistér, H. Stewénius, and E. Grossmann, "Non-parametric self-calibration," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 120–127, October 2005.

[381] Y. Nomura, M. Sagara, H. Naruse, and A. Ide, "Simple calibration algorithm for high-distortion-lens camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 1095–1099, November 1992.

[382] S. Oh and E. Hall, "Guidance of a mobile robot using an omnidirectional vision navigation system," in *Proceedings of* SPIE, MOBILE ROBOTS II, CAMBRIDGE, USA, pp. 288–300, 1987.

[383] S. Oh and E. Hall, "Calibration of an omnidirectional vision navigation system using an industrial robot," *Optical Engineering*, vol. 28, no. 9, pp. 955–962, 1989.

[384] T. Okatani and K. Deguchi, "On photometric aspects of catadioptric cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 1106–1113, 2001.

[385] J. Oliensis, "Exact two-image structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1618–1633, 2002.

[386] M. Ollis, H. Herman, and S. Singh, "Analysis and design of panoramic stereo vision using equi-angular pixel cameras," Technical Report CMU-RI-TR-99-04, Carnegie Mellon University, 1999.

[387] V. Orekhov, B. Abidi, C. Broaddus, and M. Abidi, "Universal camera calibration with automatic distortion model selection," in *Proceedings of the IEEE International Conference on Image Processing, San Antonio, USA*, pp. 397–400, 2007.

[388] R. Orghidan, J. Salvi, and E. Mouaddib, "Calibration of a structured light-based stereo catadioptric sensor," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[389] T. Pajdla, "Geometry of two-slit camera," Technical Report CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague, March 2002.

[390] T. Pajdla, "Stereo with oblique cameras," *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 161–170, 2002.

[391] T. Pajdla, T. Svoboda, and V. Hlavac, "Epipolar geometry of central panoramic cameras," in *Panoramic Vision: Sensors, Theory, and Applications*, (R. Benosman and S. Kang, eds.), pp. 85–114, Springer-Verlag, 2001.

[392] F. Pardo, B. Dierickx, and D. Scheffer, "CMOS foveated image sensor: Signal scaling and small geometry effects," *IEEE Transactions on Electron Devices*, vol. 44, no. 10, pp. 1731–1737, 1997.

[393] C. Pégard and E. Mouaddib, "A mobile robot using a panoramic view," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 89–94, April 1996.

[394] S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 1395–1401, 1999.

[395] S. Peleg, M. Ben-Ezra, and Y. Pritch, "Omnistereo: Panoramic stereo imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 279–290, March 2001.

[396] S. Peleg, Y. Pritch, and M. Ben-Ezra, "Cameras for stereo panoramic imaging," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pp. 208–214, 2000.

[397] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1144–1154, October 2000.

[398] M. Penna, "Camera calibration: A quick and easy way to determine the scale factor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 1240–1245, December 1991.

[399] C. Perwass and G. Sommer, "The inversion camera model," in *Proceedings of the 28th DAGM Symposium, Berlin, Germany*, pp. 647–656, 2006.

[400] R. Petty, S. Godber, M. Robinson, and J. Evans, "3-D vision systems using rotating 1-D sensors," in *Proceedings of the IEE Colloquium on Application of Machine Vision. London, UK*, pp. 6/1–6/6, 1995.

[401] B. Peuchot and M. Saint-André, "CCD camera calibration virtual equivalent model," in *14th Annual International Conference IEEE EMBS, Paris*, pp. 1960–1961, October 1992.

[402] R. Pless, "Using many cameras as one," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, pp. 587–593, June 2003.

[403] G. Poivilliers, "Une chambre photographique quadruple de 20 milllimètres de focale," in *Proceedings of the IV ISPRS-Congress, Paris, France*, pp. 132–134, 1934.

[404] M. Pollefeys, R. Koch, and L. van Gool, "A simple and efficient rectification method for general motion," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece*, pp. 496–501, 1999.

[405] J. Ponce, "What is a camera?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA*, 2009.

[406] V. Popescu, J. Dauble, C. Mei, and E. Sacks, "An efficient error-bounded general camera model," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece*, 2006.

[407] I. Powell, "Panoramic lens," *Applied Optics*, vol. 33, no. 31, pp. 7356–7361, 1994.

[408] B. Prescott and G. McLean, "Line-based correction of radial lens distortion," *Graphical Models and Image Processing*, vol. 59, pp. 39–47, January 1997.

[409] C. Pulfrich, "Über die stereoskopische Betrachtung eines Gegenstandes und seines Spiegelbildes," *Zeitschrift für Instrumentenkunde*, vol. 25, no. 4, pp. 93–96, 1905.

[410] M. Qiu and S. Ma, "The nonparametric approach for camera calibration," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, (E. Grimson, ed.), pp. 224–229, IEEE, IEEE Computer Society Press, June 1995.

[411] P. Rademacher and G. Bishop, "Multiple-center-of-projection images," in *Proceedings of the SIGGRAPH*, pp. 199–206, 1998.

[412] N. Ragot, "Conception d'un capteur de stéréovision omnidirectionnelle: Architecture, étalonnage et applications à la reconstruction de scènes 3D," PhD thesis, Université de Rouen, France, 2009.

[413] N. Ragot, J.-Y. Ertaud, X. Savatier, and B. Mazari, "Calibration of a panoramic stereovision sensor: Analytical vs interpolation-based methods," in *Proceedings of the 32nd Annual Conference of the IEEE Industrial Electronics Society, Paris, France*, pp. 4130–4135, 2006.

[414] S. Ramalingam, S. Lodha, and P. Sturm, "A generic structure-from-motion algorithm for cross-camera scenarios," in *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, pp. 175–186, May 2004.

[415] S. Ramalingam, S. Lodha, and P. Sturm, "A generic structure-from-motion framework," *Computer Vision and Image Understanding*, vol. 103, pp. 218–228, September 2006.

[416] S. Ramalingam and P. Sturm, "Minimal solutions for generic imaging models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, USA*, June 2008.

[417] S. Ramalingam, P. Sturm, and E. Boyer, "A factorization based self-calibration for radially symmetric cameras," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission, Chapel Hill, USA*, June 2006.

[418] S. Ramalingam, P. Sturm, and S. Lodha, "Towards complete generic camera calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 1093–1098, June 2005.

[419] S. Ramalingam, P. Sturm, and S. Lodha, "Towards generic self-calibration of central cameras," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, pp. 20–27, October 2005.

[420] S. Ramalingam, P. Sturm, and S. Lodha, "Theory and calibration algorithms for axial cameras," in *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, pp. 704–713, January 2006.

[421] S. Ramalingam, P. Sturm, and S. Lodha, "Generic self-calibration of central cameras," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 210–219, 2010.

[422] B. Ramsgaard, I. Balslev, and J. Arnspang, "Mirror-based trinocular systems in robot-vision," in *Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain*, pp. 499–502, 2000.

[423] O. Reading, "The nine lens air camera of the U.S. coast and geodetic survey," *Photogrammetric Engineering*, vol. IV, no. 3, pp. 184–192, 1938.

[424] D. Rees, "Panoramic television viewing system, U.S. patent no. 3,505,465," 1970.

[425] D. Rees, "Panoramic imaging block for three-dimensional space, U.S. patent no. 4,566,763," 1985.

[426] F. Remondino and C. Fraser, "Digital camera calibration methods: considerations and comparisons," in *ISPRS Commission V Symposium, Dresden, Germany*, pp. 266–272, 2006.

[427] S. Remy, M. Dhome, N. Daucher, and J. Lapresté, "Estimating the radial distortion of an optical system; effect on a localization process," in *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pp. 997–1001, 1994.

[428] C. Ricolfe-Viala and A.-J. Sánchez-Salmerón, "Robust metric calibration of non-linear camera lens distortion," *Pattern Recognition*, vol. 43, no. 4, pp. 1688–1699, 2010.

[429] D. Roberts, "History of lenticular and related autostereoscopic methods, white paper," Leap Technologies, 2003.

[430] R. Roelofs, "Distortion, principal point, point of symmetry and calibrated principal point," *Photogrammetria*, vol. 7, pp. 49–66, 1951.

[431] S. Roy, J. Meunier, and I. Cox, "Cylindrical rectification to minimize epipolar distortion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pp. 393–399, 1997.

[432] R. Sagawa, N. Kurita, T. Echigo, and Y. Yagi, "Compound catadioptric stereo sensor for omnidirectional object detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan*, pp. 2612–2617, September 2004.

[433] R. Sagawa, M. Takatsuji, T. Echigo, and Y. Yagi, "Calibration of lens distortion by structured-light scanning," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada*, pp. 1349–1354, August 2005.

[434] K. Sarachik, "Visual navigation: Constructing and utilizing simple maps of an indoor environment," PhD thesis, Massachusetts Institute of Technology, published as technical report AITR-1113 of the MIT AI Lab, 1989.

[435] H. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 3, pp. 235–243, 1999.

[436] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China*, pp. 5695–5701, 2006.

[437] D. Scaramuzza and R. Siegwart, "A practical toolbox for calibrating omnidirectional cameras," in *Vision Systems: Applications*, (G. Obinata and A. Dutta, eds.), pp. 297–310, I-Tech Education and Publishing, Vienna, Austria, 2007.

[438] T. Scheimpflug, "Der Perspektograph und seine Anwendung," *Photographische Korrespondenz*, 1906.

[439] D. Schneider, E. Schwalbe, and H.-G. Maas, "Validation of geometric models for fisheye lenses," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 3, pp. 259–266, 2009.

[440] H. Schreier, D. Garcia, and M. Sutton, "Advances in light microscope stereo vision," *Experimental Mechanics*, vol. 44, pp. 278–288, June 2004.

[441] E. Schwartz, "Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding," *Vision Research*, vol. 20, no. 8, pp. 645–669, 1980.

[442] G. Schweighofer and A. Pinz, "Globally optimal O(n) solution to the PnP problem for general camera models," in *Proceedings of the 19th British Machine Vision Conference, Leeds, England*, pp. 1–10, 2008.

[443] G. Schweighofer, S. Šegvić, and A. Pinz, "Online/realtime structure and motion for general camera models," in *Proceedings of the IEEE Workshop on Applications of Computer Vision, Copper Mountain, USA*, 2008.

[444] S. Seitz, A. Kalai, and H.-Y. Shum, "Omnivergent stereo," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 159–172, 2002.

[445] S. Seitz and J. Kim, "The space of all stereo images," *International Journal of Computer Vision*, vol. 48, pp. 21–38, June 2002.

[446] O. Shakernia, R. Vidal, and S. Sastry, "Omnidirectional egomotion estimation from back-projection flow," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[447] O. Shakernia, R. Vidal, and S. Sastry, "Structure from small baseline motion with central panoramic cameras," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[448] S. Shih, Y. Hung, and W. Lin, "Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem," *Optical Engineering*, vol. 32, pp. 138–149, January 1993.

[449] S. Shih, Y. Hung, and W. Lin, "When should we consider lens distortion in camera calibration," *Pattern Recognition*, vol. 28, no. 3, pp. 447–461, 1995.

[450] J. Šivic and T. Pajdla, "Geometry of concentric multiperspective panoramas," Technical Report CTU-CMP-2002-05, Center for Machine Perception, Czech Technical University, Prague, 2002.

[451] C. Slama, ed., *Manual of Photogrammetry.* Falls Church, Virginia, USA: American Society of Photogrammetry and Remote Sensing, 4th Edition, 1980.

[452] L. Smadja, R. Benosman, and J. Devars, "Cylindrical sensor calibration using lines," in *Proceedings of the IEEE International Conference on Image Processing, Singapore*, pp. 1851–1854, 2004.

[453] P. Smith, K. Johnson, and M. Abidi, "Efficient techniques for wide-angle stereo vision using surface projection models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 113–118, 1999.

[454] W. Smith, N. Vakil, and S. Maislin, "Correction of distortion in endoscope images," *IEEE Transactions on Medical Imaging*, vol. 11, pp. 117–122, March 1992.

[455] D. Southwell, A. Basu, M. Fiala, and J. Reyda, "Panoramic stereo," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 378–382, 1996.

[456] L. Spacek, "Coaxial omnidirectional stereopsis," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, pp. 354–365, Lecture Notes in Computer Science, 2004.

[457] M. Srinivasan, "New class of mirrors for wide-angle imaging," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[458] R. Steele and C. Jaynes, "Overconstrained linear estimation of radial distortion and multi-view geometry," in *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, pp. 253–264, 2006.

[459] T. Stehle, D. Truhn, T. Aach, C. Trautwein, and J. Tischendorf, "Camera calibration for fish-eye lenses in endoscopy with an application to 3D reconstruction," in *Proceedings IEEE International Symposium on Biomedical Imaging, Washington, D.C.*, pp. 1176–1179, 2007.

[460] G. Stein, "Accurate internal camera calibration using rotation, with analysis of sources of error," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, (E. Grimson, ed.), pp. 230–236, IEEE Computer Society Press, June 1995.

[461] G. Stein, "Lens distortion calibration using point correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pp. 602–608, 1997.

[462] D. Stevenson and M. Fleck, "Robot aerobics: Four easy steps to a more flexible calibration," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pp. 34–39, 1995.

[463] D. Stevenson and M. Fleck, "Nonparametric correction of distortion," in *Proceedings of the IEEE Workshop on Applications of Computer Vision, Sarasota, Florida*, pp. 214–219, 1996.

[464] H. Stewénius, D. Nistér, M. Oskarsson, and K. Åström, "Solutions to minimal generalized relative pose problems," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[465] H. Stewénius, F. Schaffalitzky, and D. Nistér, "How hard is three-view triangulation really?," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 686–693, 2005.

[466] R. Strand and E. Hayman, "Correcting radial distortion by circle fitting," in *Proceedings of the 16th British Machine Vision Conference, Oxford, England*, 2005.

[467] D. Strelow, J. Mishler, D. Koes, and S. Singh, "Precise omnidirectional camera calibration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, pp. 689–694, 2001.

[468] D. Strelow and S. Singh, "Reckless motion estimation from omnidirectional image and inertial measurements," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[469] P. Sturm, "Self-calibration of a moving zoom-lens camera by pre-calibration," *Image and Vision Computing*, vol. 15, pp. 583–589, August 1997.

[470] P. Sturm, "A method for 3D reconstruction of piecewise planar objects from single panoramic images," in *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, pp. 119–126, IEEE, June 2000.

[471] P. Sturm, "Mixing catadioptric and perspective cameras," in *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pp. 37–44, June 2002.

[472] P. Sturm, "Multi-view geometry for general camera models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 206–212, June 2005.

[473] P. Sturm and J. Barreto, "General imaging geometry for central catadioptric cameras," in *Proceedings of the 10th European Conference on Computer Vision, Marseille, France*, (D. Forsyth, P. Torr, and A. Zisserman, eds.), pp. 609–622, October 2008.

[474] P. Sturm, Z. Cheng, P. Chen, and A. Poo, "Focal length calibration from two views: Method and analysis of singular cases," *Computer Vision and Image Understanding*, vol. 99, pp. 58–95, July 2005.

[475] P. Sturm and S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pp. 432–437, June 1999.

[476] P. Sturm and L. Quan, "Affine stereo calibration," Technical Report LIFIA-29, LIFIA–IMAG, Grenoble, France, June 1995.

[477] P. Sturm and S. Ramalingam, "A generic concept for camera calibration," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, (T. Pajdla and J. Matas, eds.), pp. 1–13, Springer-Verlag, May 2004.

[478] P. Sturm, S. Ramalingam, and S. Lodha, "On calibration, structure from motion and multi-view geometry for generic camera models," in *Imaging Beyond the Pinhole Camera*, (K. Daniilidis and R. Klette, eds.), Springer-Verlag, August 2006.

[479] T. Svoboda, "Central Panoramic Cameras: Design, Geometry, Egomotion," PhD thesis, Faculty of Electrical Engineering, Czech Technical University, Prague, September 1999.

[480] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multicamera self-calibration for virtual environments," *Presence*, vol. 14, no. 4, pp. 407–422, 2005.

[481] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *International Journal of Computer Vision*, vol. 49, no. 1, pp. 23–37, 2002.

[482] T. Svoboda, T. Pajdla, and V. Hlaváč, "Epipolar geometry for panoramic cameras," in *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, pp. 218–231, 1998.

[483] R. Swaminathan, M. Grossberg, and S. Nayar, "A perspective on distortions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, pp. 594–601, June 2003.

[484] R. Swaminathan, M. Grossberg, and S. Nayar, "Non-single viewpoint catadioptric cameras: Geometry and analysis," *International Journal of Computer Vision*, vol. 66, no. 3, pp. 211–229, 2006.

[485] R. Swaminathan and S. Nayar, "Nonmetric calibration of wide-angle lenses and polycameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1172–1178, 2000.

[486] R. Swaminathan, S. Nayar, and M. Grossberg, "Designing mirrors for catadioptric systems that minimize image errors," in *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, 2004.

[487] R. Szeliski, "Image alignment and stitching: A tutorial," *Foundations and Trends in Computer Graphics and Vision*, vol. 2, pp. 1–104, December 2006.

[488] A. Takeya, T. Kuroda, K.-I. Nishiguchi, and A. Ichikawa, "Omnidirectional vision system using two mirrors," in *Proceedings of the SPIE, Novel Optical Systems and Large-Aperture Imaging*, pp. 50–60, 1998.

[489] J.-P. Tardif, P. Sturm, and S. Roy, "Self-calibration of a general radially symmetric distortion model," in *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, (H. Bischof and A. Leonardis, eds.), pp. 186–199, May 2006.

[490] J.-P. Tardif, P. Sturm, and S. Roy, "Plane-based self-calibration of radial distortion," in *Proceedings of the 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil*, IEEE Computer Society Press, October 2007.

[491] J.-P. Tardif, P. Sturm, M. Trudeau, and S. Roy, "Calibration of cameras with radially symmetric distortion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1552–1566, 2009.

[492] W. Tayman, "Analytical multicollimator camera calibration," *Photogrammetria*, vol. 34, pp. 179–197, 1978.

[493] S. Teller, "Toward urban model acquisition from geo-located images," in *Proceedings of the Pacific Conference on Computer Graphics and Applications, Singapore*, pp. 45–51, 1998.

[494] S. Teller and M. Hohmeyer, "Determining the lines through four lines," *Journal of Graphics Tools*, vol. 4, no. 3, pp. 11–22, 1999.

[495] L. Teodosio and W. Bender, "Salient video stills: Content and context preserved," in *Proceedings of the first ACM International Conference on Multimedia, Anaheim, USA*, pp. 39–46, 1993.

[496] L. Teodosio and M. Mills, "Panoramic overviews for navigating real-world scenes," in *Proceedings of the first ACM International Conference on Multimedia, Anaheim, USA*, pp. 359–364, 1993.

[497] W. Teoh and X. Zhang, "An inexpensive stereoscopic vision system for robots," in *Proceedings of the IEEE International Conference on Robotics and Automation, Atlanta, Georgia, USA*, pp. 186–189, 1984.

[498] H. Teramoto and G. Xu, "Camera calibration by a single image of balls: From conics to the absolute conic," in *Proceedings of the Fifth Asian Conference on Computer Vision, Melbourne, Australia*, pp. 499–506, 2002.

[499] R. Thiele, "Métrophotographie aérienne à l'aide de mon autopanoramographe," *International Archives of Photogrammetry*, vol. 1, no. 1, pp. 35–45, 1908.

[500] S. Thirthala and M. Pollefeys, "Multi-view geometry of 1D radial cameras and its application to omnidirectional camera calibration," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 1539–1546, October 2005.

[501] S. Thirthala and M. Pollefeys, "The radial trifocal tensor: A tool for calibrating the radial distortion of wide-angle cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA*, pp. 321–328, 2005.

[502] E. Thompson, "The seven lens air survey camera," *Photogrammetric Engineering*, vol. IV, no. 3, pp. 137–145, 1938.

[503] T. Thormählen, H. Broszio, and I. Wassermann, "Robust line-based calibration of lens distortion from a single view," in *Proceedings of the MIRAGE Conference on Computer Vision/Computer Graphics Collaboration*

for Model-based Imaging, Rendering, Image Analysis and Graphical Special Effects, Rocquencourt, France, pp. 105–112, 2003.

[504] G. Tissandier, *La photographie en ballon.* Gauthier-Villars, 1886.

[505] M. Tistarelli and G. Sandini, "On the advantage of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 401–410, 1993.

[506] C. Toepfer and T. Ehlgen, "A unifying omnidirectional camera model and its applications," in *Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Rio de Janeiro, Brazil*, 2007.

[507] A. Tolvanen, C. Perwass, and G. Sommer, "Projective model for central catadioptric cameras using Clifford algebra," in *Proceedings of the 27th DAGM Symposium, Vienna, Austria*, pp. 192–199, 2005.

[508] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision*, vol. 9, pp. 137–154, November 1992.

[509] A. Torii, A. Sugimoto, and A. Imiya, "Mathematics of a multiple omnidirectional system," in *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, Wisconsin, USA*, 2003.

[510] J. Torres and J. Menéndez, "A practical algorithm to correct geometrical distortion of image acquisition cameras," in *Proceedings of the IEEE International Conference on Image Processing, Singapore*, pp. 2451–2454, 2004.

[511] B. Triggs, "Matching constraints and the joint image," in *Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, Massachusetts, USA*, (E. Grimson, ed.), pp. 338–343, IEEE, IEEE Computer Society Press, June 1995.

[512] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment — a modern synthesis," in *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, Corfu, Greece*, (B. Triggs, A. Zisserman, and R. Szeliski, eds.), pp. 298–372, Springer-Verlag, 2000.

[513] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall, 1998.

[514] R. Tsai, "An efficient and accurate camera calibration technique for 3D machine vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA*, pp. 364–374, 1986.

[515] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, pp. 323–344, August 1987.

[516] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy, "Using plane + parallax for calibrating dense camera arrays," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA*, pp. 2–9, 2004.

[517] B. Vandeportaele, "Contributions à la vision omnidirectionnelle: Étude, Conception et Étalonnage de capteurs pour l'acquisition d'images et la modélisation 3D," PhD thesis, Institut National Polytechnique de Toulouse, France, in french, December 2006.

[518] B. Vandeportaele, M. Cattoen, P. Marthon, and P. Gurdjos, "A new linear calibration method for paracatadioptric cameras," in *Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong*, pp. 647–651, 2006.

[519] P. Vasseur and E. Mouaddib, "Central catadioptric line detection," in *Proceedings of the 15th British Machine Vision Conference, Kingston upon Thames, England*, 2004.

[520] A. Wang, T. Qiu, and L. Shao, "A simple method of radial distortion correction with centre of distortion estimation," *Journal of Mathematical Imaging and Vision*, vol. 35, no. 3, pp. 165–172, 2009.

[521] J. Wang, F. Shi, J. Zhang, and Y. Liu, "A new calibration model of camera lens distortion," *Pattern Recognition*, vol. 41, no. 2, pp. 607–615, 2008.

[522] G. Wei and S. Ma, "Two plane camera calibration: a unified model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pp. 133–138, 1991.

[523] G. Wei and S. Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 469–480, 1994.

[524] G.-Q. Wei and S. Ma, "A complete two-plane camera calibration method and experimental comparisons," in *Proceedings of the 4th IEEE International Conference on Computer Vision, Berlin, Germany*, pp. 439–446, 1993.

[525] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accurate evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 965–980, October 1992.

[526] C. Wheatstone, "Contributions to the physiology of vision — part the first — on some remarkable, and hitherto unobserved, phenomena of binocular vision," *Philosophical Transactions of the Royal Society of London*, vol. 128, pp. 371–394, 1838.

[527] C. Wheatstone, "Contributions to the physiology of vision — part the second — on some remarkable, and hitherto unobserved, phenomena of binocular vision (continued)," *Philosophical Transactions of the Royal Society of London*, vol. 142, pp. 1–17, 1852.

[528] Wikipedia, "Catadioptric system," http://en.wikipedia.org/wiki/Catadioptric_system.

[529] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA*, pp. 294–301, 2004.

[530] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 765–776, 2005.

[531] A. Wiley and K. Wong, "Metric aspects of zoom vision," *International Archives of Photogrammetry and Remote Sensing*, vol. 28, no. 5, pp. 112–118, 1990. also in SPIE Vol. 1395: Close-Range Photogrammetry Meets Machine Vision (1990).

[532] R. Willson and S. Shafer, "A perspective projection camera model for zoom lenses," in *Proceedings of the Second Conference on Optical 3D Measurement Techniques, Zürich, Switzerland*, October 1993.

[533] R. Willson and S. Shafer, "What is the center of the image?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pp. 670–671, 1993.

[534] K. Wong, "Photogrammetric calibration of television systems," in *Proceedings of the XII ISPRS-Congress, Ottawa, Canada*, 1972.

[535] D.-M. Woo and D.-C. Park, "Implicit camera calibration based on a nonlinear modeling function of an artificial neural network," in *Proceedings of the 6th International Symposium on Neural Networks, Wuhan, China*, pp. 967–975, Springer-Verlag, 2009.

[536] R. Wood, "Fish-eye views, and vision under water," *Philosophical Magazine Series 6*, vol. 12, no. 68, pp. 159–162, 1906.

[537] F. Wu, F. Duan, Z. Hu, and Y. Wu, "A new linear algorithm for calibrating central catadioptric cameras," *Pattern Recognition*, vol. 41, no. 10, pp. 3166–3172, 2008.

[538] A. Würz-Wessel, "Free-formed Surface Mirrors in Computer Vision Systems," PhD thesis, Eberhard-Karls-Universität Tübingen, Germany, June 2003.

[539] Y. Xiong and K. Turkowski, "Creating image-based VR using a self-calibrating fisheye lens," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pp. 237–243, June 1997.

[540] M. Yachida, "Omnidirectional sensing and combined multiple sensing," in *Proceedings of the IEEE and ATR Workshop on Computer Vision for Virtual Reality Based Human Communications, Bombay, India*, pp. 20–27, 1998.

[541] Y. Yagi, "Omnidirectional sensing and applications," *IEICE Transactions on Information and Systems*, vol. E82-D, no. 3, pp. 568–579, 1999.

[542] Y. Yagi and S. Kawato, "Panoramic scene analysis with conic projection," in *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems, Ibaraki, Japan*, pp. 181–187, 1990.

[543] Y. Yagi, W. Nishii, K. Yamazawa, and M. Yachida, "Rolling motion estimation for mobile robot by using omnidirectional image sensor hyperomnivision," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 946–950, 1996.

[544] Y. Yagi, H. Okumura, and M. Yachida, "Multiple visual sensing system for mobile robot," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1679–1684, 1994.

[545] K. Yamazawa, Y. Yagi, and M. Yachida, "3D line segment reconstruction by using hyperomni vision and omnidirectional hough transforming," in *Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain*, pp. 483–486, 2000.

[546] K. Yamazawa, Y. Yagi, and M. Yachida, "Omindirectional imaging with hyperboloidal projection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan*, pp. 1029–1034, 1993.

[547] S. Yelick, "Anamorphic image processing," Bachelor Thesis, Massachusetts Institute of Technology, May 1980.

[548] S. Yi and N. Ahuja, "An omnidirectional stereo vision system using a single camera," in *Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong*, pp. 861–865, 2006.

[549] W. Yin and T. Boult, "Physical panoramic pyramid and noise sensitivity in pyramids," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pp. 90–97, 2000.

[550] X. Ying and Z. Hu, "Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, pp. 442–355, 2004.

[551] X. Ying and Z. Hu, "Catadioptric camera calibration using geometric invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1260–1271, October 2004.

[552] X. Ying and Z. Hu, "Distortion correction of fisheye lenses using a non-parametric imaging model," in *Proceedings of the Asian Conference on Computer Vision, Jeju Island, Korea*, pp. 527–532, 2004.

[553] X. Ying, Z. Hu, and H. Zha, "Fisheye lenses calibration using straight-line spherical perspective projection constraint," in *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, pp. 61–70, 2006.

[554] X. Ying and H. Zha, "Linear catadioptric camera calibration from sphere images," in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.

[555] X. Ying and H. Zha, "Geometric interpretations of the relation between the image of the absolute conic and sphere images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2031–2036, December 2006.

[556] X. Ying and H. Zha, "Identical projective geometric properties of central catadioptric line images and sphere images with applications to calibration," *International Journal of Computer Vision*, vol. 78, pp. 89–105, June 2008.

[557] N. Yokobori, P. Yeh, and A. Rosenfeld, "Selective geometric correction of images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA*, pp. 530–533, 1986.

[558] J. Yu, Y. Ding, and L. McMillan, "Multiperspective modeling and rendering using general linear cameras," *Communications in Information and Systems*, vol. 7, no. 4, pp. 359–384, 2007.

[559] J. Yu and L. McMillan, "General linear cameras," in *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, (T. Pajdla and J. Matas, eds.), pp. 14–27, Springer-Verlag, May 2004.

[560] J. Yu and L. McMillan, "Multiperspective projection and collineation," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 580–587, 2005.

[561] W. Yu, "Image-based lens geometric distortion correction using minimization of average bicoherence index," *Pattern Recognition*, vol. 37, no. 6, pp. 1175–1187, 2004.

[562] K. Zaar, "Spiegelphotographie und ihre Auswertung zu Messzwecken," *International Archives of Photogrammetry*, vol. 3, no. 2, pp. 96–105, 1912.

[563] K. Zaar, "Beiträge zur Spiegelphotogrammetrie," *International Archives of Photogrammetry*, vol. 3, no. 4, pp. 269–276, 1913.

[564] L. Zelnik-Manor, G. Peters, and P. Perona, "Squaring the circle in panoramas," in *Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China*, pp. 1292–1299, 2005.

[565] C. Zhang, J. Helferty, G. McLennan, and W. Higgins, "Nonlinear distortion correction in endoscopic video images," in *Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada*, pp. 439–442, 2000.

[566] Z. Zhang, "On the epipolar geometry between two images with lens distortion," in *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, pp. 407–411, IEEE Computer Society Press, August 1996.

[567] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[568] Z. Zhang, "Camera calibration with one-dimensional objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 892–899, 2004.

[569] Z. Zhang and H. Tsui, "3D reconstruction from a single view of an object and its image in a plane mirror," in *Proceedings of the 14th International Conference on Pattern Recognition, Brisbane, Australia*, pp. 1174–1176, 1998.

[570] J. Zheng and S. Tsuji, "From anorthoscope perception to dynamic vision," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1154–1160, 1990.

[571] J. Zheng and S. Tsuji, "Panoramic representation of scenes for route understanding," in *Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, USA*, pp. 161–167, 1990.

[572] J. Zheng and S. Tsuji, "Panoramic representation for route recognition by a mobile robot," *International Journal of Computer Vision*, vol. 9, no. 1, pp. 55–76, 1992.

[573] Z. Zhu, "Omnidirectional stereo vision," in *Proceedings of the Workshop on Omnidirectional Vision, Budapest, Hungary*, 2001.

[574] Z. Zhu, E. Riseman, and A. Hanson, "Geometrical modeling and real-time vision applications of a panoramic annular lens (PAL) camera system," Technical Report 99-11, University of Massachusetts at Amherst, 1999.

[575] H. Zollner and R. Sablatnig, "A method for determining geometrical distortion of off-the-shelf wide-angle cameras," in *Proceedings of the DAGM Symposium on Pattern Recognition, Vienna, Austria*, pp. 224–229, 2005.

[576] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall, "Mosaicing new views: The crossed-slit projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 741–754, June 2003.

[577] A. Zou, Z. Hou, L. Zhang, and M. Tan, "A neural network-based camera calibration method for mobile robot localization problems," in *Proceedings of the Second International Symposium on Neural Networks, Chongqing, China*, pp. 277–284, Springer-Verlag, 2005.