Two-View Geometry of Omnidirectional Cameras

A Dissertation Presented to the Faculty of the Electrical Engineering of the Czech Technical University in Prague in Partial Fulfilment of the Requirements for the Ph.D. Degree in Study Programme No. P2612 - Electrotechnics and Informatics, branch No. 3902V035 - Artificial Intelligence and Biocybernetics, by

Branislav Mičušík

June 17, 2004

Thesis Advisor

Prof. Ing. Václav Hlaváč, CSc.

Thesis Co-Advisor

Ing. Tomáš Pajdla, Ph.D.

Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo náměstí 13, 121 35 Prague 2, Czech Republic
fax: +420 224 357 385, phone: +420 224 357 465
http://cmp.felk.cvut.cz

Abstract

Conventional cameras in computer vision usually stand for central cameras with restricted angle of view. There are cameras, named omnidirectional, with field of view above 180°, and thus the ability to to see very large part of surrounding scene at one instant. In this thesis, we deal with the omnidirectional cameras, their geometry and properties in order to build a 3D metric reconstruction from uncalibrated omnidirectional images.

Computer vision is concerned with the problem of building a 3D reconstruction from two or many 2D images. The camera calibration is a one of fundamental problems in the procedure from 2D images to 3D reconstruction. Often, no information about mirror or lens parameters and no calibration object are available. In such situations, it is still often easy to acquire two or more images of the surrounding scene. Point correspondences can be established manually or automatically and the auto-calibration can be performed from point matches only. This idea motivates our work. The main contribution of the thesis is that we provide a technique for *auto-calibrating the omnidirectional cameras and building a 3D metric reconstruction automatically from point correspondences (possibly) contaminated by mismatches.*

The theory of central omnidirectional cameras is presented. The emphasis is put on geometry. The standard projection model is extended in order to describe omnidirectional camera geometry. An omnidirectional image formation is studied and used in a formulation of the camera auto-calibration. The proposed auto-calibration technique can be viewed as a generalization of the Fitzgibbon's method for radial distortion estimation. The robust auto-calibration method leading to the Polynomial Eigenvalue Problem is derived for catadioptric (parabolic, hyperbolic, spherical mirror) and for dioptric (fish-eye lenses) cameras. The performance on real data is demonstrated. Experiments show that the central model for real catadioptric cameras is often not sufficient to obtain accurate 3D reconstruction but it is always sufficient enough to get an initial estimate of the camera model and motion. The real catadioptric cameras have to be treated as non-central to improve the accuracy of Structure from Motion estimation.

It follows from the thesis, that the complex problem - *Structure from Motion from real uncalibrated panoramic cameras* - should be solved by a hierarchical approach using appropriate approximations (a linearization of a camera model or a substitution of the central model for the non-central one) in each step of the hierarchy. Solving the problem of Structure from Motion directly by using the most precise model is usually too complex and thus very unstable or computationally expensive. The thesis suggests a way how to build such a hierarchy for many types of central and slightly non-central omnidirectional cameras and how to compute precise 3D metric reconstruction from omnidirectional images.

Acknowledgement

I would like to express my thanks to Prof. Václav Hlaváč, the head of the Center for Machine Perception (CMP), without whom I would never have visited and have become a member of the CMP. He deserves my deepest gratitude for adopting me into the group, for providing excellent research and social facilities. I feel very privileged to have had an opportunity to study and work in this excellent research group.

I am greatly indebted to my co-advisor Dr. Tomáš Pajdla for introducing me to the world of computer vision. He has skillfully guided me throughout my research with his supportive attitude, vast scientific knowledge and capacity. His continuous support, patience, willingness, considerable help and guidance has been of outmost importance for finishing my PhD study.

The work would be impossible or at least much less enjoyable to complete without the support, assistance and discussions with my colleagues from the CMP. Eva Matysková, Martina Budošová and Daniel Večerka significantly helped me with administration and technical stuff. Dr. Tomáš Pajdla and Prof. Václav Hlaváč are greatly acknowledged for their careful revision, valuable suggestions, and the useful and thoughtful comments on the manuscript of this thesis.

Finally, my deep thanks belong to my fiancée Martina and my family for all their support, tireless encouragement, inspiration and motivation during these years.

I gratefully acknowledge grants of the Grant Agency of the Czech Technical University CTU 0209513, CTU 0306813, the grant of the Grant Agency of the Czech Republic GACR 102/01/0971, the grants of the Czech Ministry of Education MSM 212300013, MSMT Kontakt 22-2003-04, and the European Union projects EU FP5 OMNIVIEWS IST-1999-29017 and EU FP5 BeNoGo IST-2001-39184 which supported my research.

Contents

Keywords: computer vision, stereo vision, uncalibrated omnidirectional camera, auto-calibration, epipolar geometry, stereo geometry, 3D reconstruction

1 Introduction

The cameras map surrounding space through optical systems (lenses, mirrors, filters, etc.) onto photo-sensitive devices (e.g., film, CCD, CMOS sensors). Various combinations of optical elements lead to various types of cameras, to standard or non-standard ones. Sometimes it is an exigency arised from real world and sometimes just a challenge of intellectual curiosity to understand the geometry of various types of cameras.

The camera can be understood as a ray-based sensing device. We say that a camera is central, or has a single viewpoint, or a projection center, if all rays intersect in a single point. The rays may be completely independent for an arbitrary camera, in particular, they do not have to intersect in a single point. With respect to whether the rays intersect or not, cameras can be classified as *central* and *non-central* ones [71]. Depending on the field of view of a camera, or in another words, how large part of the surrounding world can be observed by the camera, the cameras divide into *directional* and *omnidirectional* ones.

Our interest is focused on *central omnidirectional* cameras which cannot be described by the standard perspective model [26, 43]. See Figure 1.1 for examples of central catadioptric (combination of lenses and mirrors) and dioptric (only lenses) omnidirectional cameras. The reason why a center of projection is so desirable is that the models, and therefore algorithms, are simpler and sometimes it is the need to be able to generate pure perspective images from the sensed ones. The advantage of omnidirectional cameras is primarily that they have a large field of view and thus the ability to see a large part of the surrounding scene in one instant. The large angle of view often allows to establish more spacious point correspondences which leads to a more complete 3D reconstruction from fewer images. An occurrence of degenerate scenes (e.g., when only a single plane is observed in the image) is less likely with omnidirectional images and therefore more stable ego-motion estimation is often achieved. Omnidirectional cameras found, among others, many applications in surveillance, navigation, robotics, virtual reality, and telepresence.

This thesis deals with uncalibrated central and slightly non-central omnidirectional cameras. The thesis studies omnidirectional camera geometry, properties, image formation and possibility of their auto-calibration from point correspondences in order to obtain a two-view Structure from Motion estimation. As the main contribution, the robust auto-calibration method allowing to estimate two-view geometry of omnidirectional cameras from stereo image correspondences only is developed.

Figure 1.1.: Central omnidirectional cameras and acquired images. From the top: perspective camera + hyperbolic mirror; orthographic camera + parabolic mirror; Nikon Coolpix digital camera + Nikon FC-E8 fish-eye lens; Canon EOS-1Ds + Sigma 8mm-f4-EX fish-eye lens.

1.1. Motivation

This work was motivated by desire of using using uncalibrated omnidirectional cameras, like in Figure 1.1, in stereo geometry in a similar manner as conventional directional cameras can be used, see Figure 1.2. The geometry of the central omnidirectional cameras is well understood [54]. However, *no* method allowing to robustly auto-calibrate omnidirectional cameras just from real point correspondences corrupted by mismatches was known. Therefore, our work concentrates on the uncalibrated case, where no assumptions about the scene, except for rigidity, were made (e.g., presence of lines or a calibration object in scene) to calibrate an omnidirectional camera. Often, no information about mirror or lens parameters is available as well. In such situations, however, it is still often easy to acquire two or more images of the surrounding scene and to establish point correspondences either manually or automatically.

Our ultimate goal is to design a method allowing an auto-calibration from automatically detected point correspondences between two omnidirectional images only and to

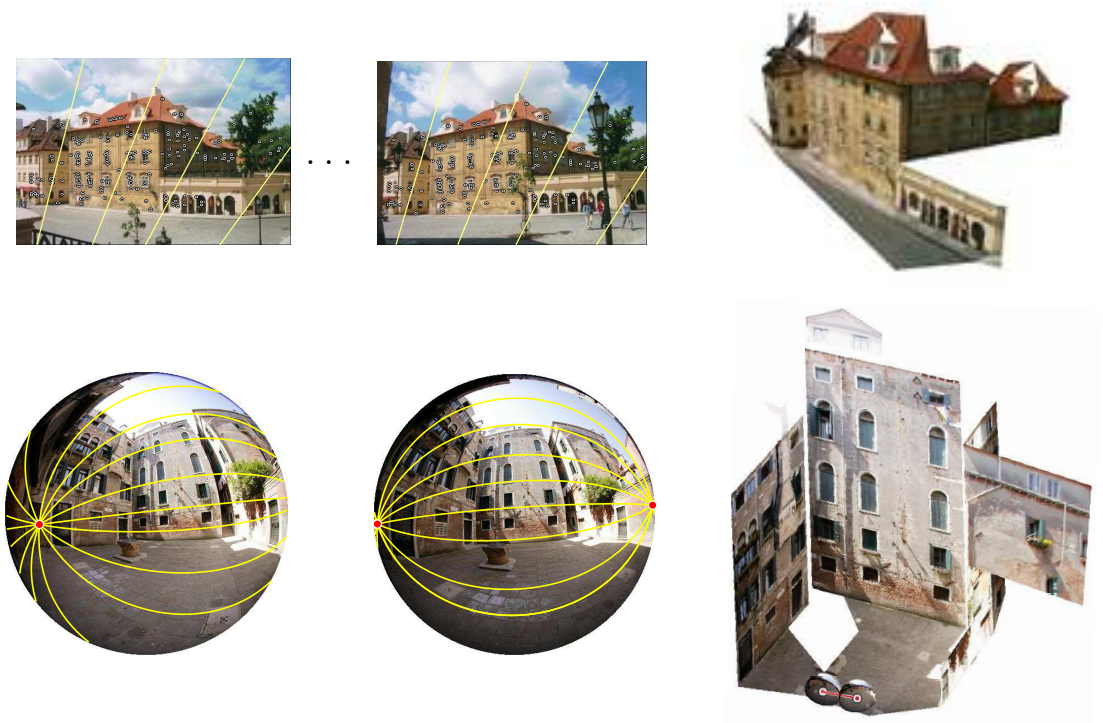


Figure 1.2.: 3D metric reconstructions. First row: The 3D metric reconstruction obtained from *many* uncalibrated conventional images. Courtesy of T. Werner et al. [95]. Second row: The same result, the 3D metric reconstruction, however, built from *two* uncalibrated omnidirectional images.

build a 3D metric (sometimes called similarity or Euclidean) reconstruction from uncalibrated omnidirectional images as Figure 1.2 (bottom) shows.

1.2. Goals of the thesis

The goal of the thesis is to develop a robust estimation of two-view geometry of uncalibrated central and slightly non-central omnidirectional cameras for the purpose of building 3D metric reconstruction from uncalibrated panoramic images. Particular goals are:

1. to find appropriate *image formation* models and
2. corresponding *camera auto-calibration* algorithms which
3. can easily be incorporated into a *robust method* handling real data for an interesting class of *central omnidirectional* (dioptric and catadioptric) cameras;
4. to find *ambiguities* in camera auto-calibration estimation affecting the 3D reconstruction, and
5. to use the auto-calibration method designed for central omnidirectional cameras for real slightly *non-central* catadioptric cameras that can be well approximated by central models.

1.3. The method

The method used in the thesis is the following. First, the theory of central omnidirectional cameras with emphasis on ambiguities in camera calibration is dealt with. Second, the auto-calibration methods are proposed for various catadioptric and dioptric cameras individually and verified on real data. Finally, as the result followed from experiments, the real catadioptric cameras are analyzed and treated as non-central cameras what leads to a more precise 3D reconstruction.

1.4. Contribution of the thesis

The main contribution of the thesis is in showing that a robust auto-calibration method based only on point correspondences and epipolar geometry works for a large class of central omnidirectional cameras. The following contributions are brought by the thesis:

1. *A robust auto-calibration method*, bases on RANSAC, for central omnidirectional cameras from point correspondences only is proposed and verified. This is important, since *no* lines or calibration objects are further needed to calibrate omnidirectional cameras.
2. *A unified model* for axially symmetric image formation functions is proposed.

3. *General conditions*, under which such a model can be auto-calibrated from stereo correspondences are given.

4. Important practical cases of catadioptric and dioptric cameras are studied and it is shown that they *all* can be successfully auto-calibrated by solving a Polynomial Eigenvalue Problem and that a reasonable Euclidean 3D reconstruction can be recovered.

5. *The real non-central catadioptric cameras* (with parabolic, hyperbolic, spherical mirror) are dealt with and their non-central models are derived. We show that the hierarchical approach (starting with approximate central model and ending with precise non-central one) should be used to solve a complex problem such as 3D metric reconstruction from two uncalibrated real (non-central) catadioptric cameras.

1.5. Structure of the thesis

The basic terminology used throughout the thesis is explained in Chapter 2. Previous and related work is reviewed in Chapter 3. Chapter 4 deals with the theory of omnidirectional camera geometry. In this chapter, an omnidirectional camera model, an image formation, camera calibration steps and epipolar geometry are described and a method for 3D reconstruction for omnidirectional cameras is shown. Chapter 5 begins with explanation of Fitzgibbon's method for radial distortion and epipolar geometry estimation. The method, designed for conventional cameras, is generalized to omnidirectional cameras and demonstrated for some types of omnidirectional cameras. The robust method based on RANSAC paradigm is suggested and some degenerate configurations are discussed. Chapter 6 focuses on real catadioptric cameras which are slightly non-central. The non-central models are derived for some types of catadioptric cameras, a method for 3D reconstruction is suggested and experimental results are shown. Finally, Chapter 7 concludes the thesis.

1.6. Authorship

I hereby certify that the results presented in this thesis were achieved during my own research in cooperation with my thesis co-advisor Tomáš Pajdla, published in [32, 62, 63, 64, 60, 66, 65], with Daniel Martinec, published in [60], and with Stefan Gächter, published in [32].

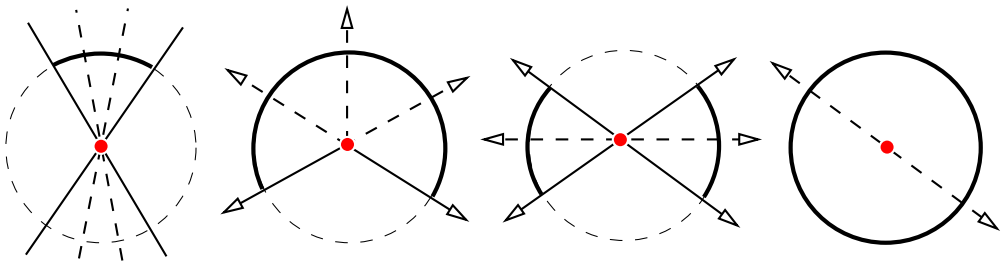
2 Notation and Concepts

In this thesis, a *camera* means a subset of the set of lines in $\mathbb{P}^3$ [72]. A subset of lines in $\mathbb{P}^3$ passing through a single point, a *projection center*, is called the *central camera*. If there is no constraint on lines of the camera then the camera is, in general, *non-central*.

We say that the camera is *directional*, see Figure 2.1(a), if its field of view is a proper subset of a hemisphere on the view-sphere [85]. For the directional camera, there exist a plane which does not contain any camera ray, hence the camera is pointed into the direction given by the normal of that plane. In another words, there exists a halfspace, including its border plane, in which no points can be seen.

Omnidirectional cameras have, depending on its construction, various fields of view, see Figure 2.1(b, c, d). *Fish-eye lenses*, with field of view larger or equal to 180°, as Figure 2.1(b) shows, are omnidirectional cameras as well. The omnidirectional cameras consisting of curved mirrors [15], called *catadioptric cameras*, have typically field of view as shown in Figure 2.1(c). We say that the camera is *completely* or *ideally omnidirectional* if it has complete field of view and its image covers the whole sphere, see Figure 2.1(c). The camera providing images covering the whole view-sphere would be the camera with no self-occlusion. However, it is difficult to realize it because some part of scene is often occluded by an image sensor.

By *the camera calibration*, we will understand the process after which we are able to compute the direction (for non-central cameras also the location) of a 3D ray for each point in a digital image. *The auto-calibration* means the calibration of the camera from



(a) (b) (c) (d)

Figure 2.1.: Type of central cameras. (a) Directional camera. The points are represented by straight lines. (b), (c), (d) Omnidirectional cameras. (b) Fish-eye lens. (c) Catadioptric camera. (d) Ideal omnidirectional camera with full field of view.

digital images without using calibration objects or assuming a special type of camera motion.

Points in two or many images are in *correspondence* or we say that they are point *matches* if they arise as projections of the same scene point. A *mismatch* or an *outlier* is a correspondence related to no real or a bad 3D scene point. *Inlier* stands for a correct match. Usually, the correspondences established by automatic methods, called *tentative* or *putative correspondences*, include both inliers and outliers. The process of establishing the tentative point matches is called *matching*. The process of detecting inliers/outliers in tentative correspondences is called *validation*.

By *the 3D reconstruction* we will understand the process of finding points in 3D space from their images in cameras. *The 3D reconstruction* also means the result of the reconstruction process, i.e., a cloud of 3D points together with camera centers w.r.t. a world coordinate system usually placed in one of the cameras. A 3D *metric* (*similarity* or *Euclidean*) reconstruction is a 3D reconstruction with correctly recovered angles [43]. The estimation of a trajectory of a moving camera together with 3D reconstruction of surrounding scene corresponds to the *Structure from Motion (SFM)* estimation.

The term *degenerate configuration* will be used to denote the configuration of cameras and point correspondences where we are not able to auto-calibrate the camera or to build a 3D reconstruction uniquely.

Used fonts in symbols

$a, b, \ldots, A, B, \ldots$	scalars
$\mathbf{a}, \mathbf{b}, \ldots, \mathbf{A}, \mathbf{B}, \ldots$	vectors
$\mathtt{A}, \mathtt{B}, \ldots$	matrices
$\mathbb{A}^n, (\mathbb{R}^n, \mathbb{P}^n), \ldots$	n-dimensional spaces (real, projective)
$\mathcal{A}, \ldots$	3D reconstruction

3 The State of the Art

There exists a number of works dealing with design, theory and applications of omnidirectional cameras. The full historical perspective and state-of-the-art in omnidirectional vision can be found in Benosman and Kang's book [10]. We will mention only works that have introduced important concept relevant to our work. We will give a quick overview of the state-of-the-art together with our contribution in comparison to it.

The overview of the state-of-the-art will be done in three steps. We begin with work on omnidirectional camera geometry. We will continue with omnidirectional camera calibration and finish with work on the 3D reconstruction from omnidirectional images.

3.1. Omnidirectional camera geometry

The concept of central catadioptric cameras appeared already in the presentation of René Descartes in 1637 in *Discours de la Methode* [21]. He showed that refractive as well as reflective 'ovals' (conical lenses and mirrors) focus light into a single point if they are illuminated from another properly chosen point. The idea was later re-phrased, e.g, by Feynman et al. in 1963 [28] or Hecht and Zajac in 1974 [45], and popularized into a modern language and introduced to computer vision community in 1997 by Baker and Nayar [67, 2].

There is an advantage of the omnidirectional cameras over the traditional ones in the sense of structure from motion estimation. Traditional cameras suffer from the problem that the direction of translation may lie outside of the field of view. Brodsky, Fermüller and Aloimonos [13, 27, 14] showed in 1996 that in a hemispherical image there is enough information to determine the 3D rotational and 3D translational motion up to a multiplicative constant. Svoboda et al. [87] and Gluckman and Nayar [40] showed in 1998 that restricted angle of view makes the computation of camera motion sensitive to noise. The omnidirectional cameras with large field of view suppress this problem and increase the stability of ego-motion estimation. It follows that the omnidirectional cameras provide, due to a large field of view, more stable ego-motion estimation and more complete 3D reconstruction from fewer images than standard directional perspective cameras.

In this thesis, we concentrate on *central* omnidirectional cameras which possess a single viewpoint and on real catadioptric cameras which are *slightly non-central* but can be approximated by central models.

There exist many works dealing with the theory of central omnidirectional catadioptric cameras [86, 2, 3, 35, 5, 10, 85, 84, 39, 98].

Svoboda et al. [86, 85] first introduced in 1998 the concept of epipolar geometry to central catadioptric cameras. They showed that the epipolar constraint for central catadioptric cameras holds for 3D ray direction vectors corresponding to image points. They also proved that epipolar lines (known from perspective images) are replaced by conics since the epipolar planes are projected to an image plane as conics.

As it was mentioned above, Baker and Nayar [2, 3] in 1998 popularized [21] and presented the complete class of catadioptric sensors that have a single viewpoint. They described solutions in detail, including the degenerate ones. They derived an expression for the spatial resolution of a catadioptric sensor in terms of the resolution of the cameras used to construct it, including detailed analysis of the defocus blur caused by the use of a curved mirror in a catadioptric sensor.

Geyer and Daniilidis [35] showed and proved that every catadioptric (parabolic, hyperbolic, elliptical), stereographic and standard perspective projection is equivalent to perspective mappings from the sphere to a plane with a projection center on the perpendicular to the plane passing through the center of the sphere. They derived the constraints on two projections of a sphere to be dual to each other, i.e., when antipodal points pairs of one projection are the foci of line images in another projection. They presented some invariants (e.g., the angle between two great circles) in parabolic projections obtained as a consequence that parabolic projection is equivalent to the stereographic projection and dual to the perspective one. Barreto and Araujo [5] used Geyer and Daniilidis' unified model for catadioptric cameras to study issues in the projection of lines. They derived the equations and geometric properties of general catadioptric imaging of lines. In [7], they proposed a method for fitting circles on para-catadioptric line images. Later, Geyer and Daniilidis [39] encoded the nonlinearity of the para-catadioptric projection model in a bilinear form by lifting coordinates to 4-dimensional space. It allows to compute catadioptric fundamental matrix using 15-point RANSAC without knowledge of the image of the mirror optical axis. However, the method works only for para-catadioptric cameras and it is very sensitive to noise. Ying and Hu [98] indicated that the unified imaging model (by Geyer and Daniilidis [35]) can be used for some types of fisheye lenses. The approximation of a fisheye lens model by a catadioptric one is usually possible, however, with limited accuracy only. This approach differs from our approach, since we can model any fisheye lens with any precision. Corochano and Franco presented in [8] an extension and improvement of the unified imaging model (by Geyer and Daniilidis [35]) using the conformal geometry algebra. Using this mathematical system, the analysis of diverse catadioptric mirrors becomes transparent and computationally simpler. As a result, the algebraic burden is reduced, allowing the user to work in a much more effective framework for the development of algorithms for omnidirectional vision.

Relations that exist between multiple views of a static scene, where the views can be taken by any mixture of para-catadioptric, perspective or affine cameras, were described

by Sturm [84]. The usage of this theory for motion estimation, 3D reconstruction or (self-)calibration was indicated.

In our work, we proposed a camera model suitable for both, dioptric and catadioptric cameras. Our representation includes catadioptric models introduced in [35] as well. We formulated [62] an image formation and camera calibration steps for central dioptric omnidirectional cameras, gave the projection model and proposed the method for camera auto-calibration from point correspondences only. Let us emphasize that there was no such a method before allowing to auto-calibrate omnidirectional cameras and validate the point correspondences in a one procedure.

When an imaging system does not maintain a single viewpoint, then a caustic, i.e., a locus of viewpoints in three dimensions, is formed and the system has to be treated as a non-central one. Previous work related to the non-central cameras can be found in [20, 42, 88, 82, 75, 11, 69].

Derrien and Konolige [20] approximated the non-central spherical catadioptric camera by a camera with a single viewpoint (the top of the mirror) to construct perspective images. In our work [65], we also used a central approximate model of the spherical catadioptric camera but just for obtaining an initial estimate for further bundle adjustment with the non-central model. We chose the position of the fictive center as that which minimizes the sum of squares of angular differences between approximated rays and the correct ones.

In most relevant papers by Swaminathan, Grossberg and Nayar [88, 42], the non-central devices like curved mirrors, meniscus lenses, camera clusters, or compound cameras were represented by points on the caustic and directional vectors towards scene points. Their models capture geometrical, radiometrical, and optical properties of the omnidirectional camera. Our approach to modelling the optics differs in that we represent the non-central catadioptric camera by points on the mirror surface (not on the caustic). Our model, presented in [65], captures the geometric (not radiometric and optical) properties of the camera what is sufficient for doing a 3D reconstruction. Since the caustic does not have to be computed, the derivation of our model is simpler and the computation of the Jacobian, as in [88], is not necessary. Moreover, our model can be estimated from point correspondences only.

Similar idea to ours [65] (to model slightly non-central catadioptric cameras) was suggested by Strelow et al. [82]. Their approach differs from ours by that the parameters of a camera model are obtained in calibration process assuming known 3D positions of target points.

Pless [75] illustrated how to consider a network of cameras as a single generalization in a framework proposed in [42]. He analyzed structure from motion algorithms for such type of non-central cameras and derived constraints on the optimal design of panoramic imaging system constructed from multiple cameras.

Neumann et al. [69] studied the non-central cameras and developed a new methodology for the design of eyes which interprets camera assemblies as sampling operators in the space of light rays. It allows to develop mathematical criteria for optimal eye design,

which in turn enables to build the best eye for a given task without the trial and error phase of natural evolution. A similar idea appeared in works by Hicks et al. [46, 47], in our work [32], in work by Gaspar et al. [33] and Swaminathan et al. [89] where the methods for designing a shape of a mirror to obtain special properties of the omnidirectional camera projection were suggested. Our representation and structure from motion algorithm [65] can also be applied for such type of non-central mirrors.

Bonford and Sturm [11] represented the specular surfaces by voxels in 3D space. Similar work was done by Salvarese et al.[77]. Their problem, i.e., the estimation of a specular shape from known 3D structure, is inverse to the problem solved in our work [65].

3.2. Omnidirectional camera calibration

Previous work on the estimation of camera models with lens distortion divides into two basic groups. The first one includes methods which use some knowledge about observed scene. There are methods using calibration patterns [78, 9, 4] and plumb line methods for classical [22, 99, 74] and for omnidirectional [34, 90, 12, 6, 37, 97] cameras. The second group covers methods which do not use any knowledge about the scene. There are calibration methods for standard [81, 25, 30] and for omnidirectional [96, 54, 36, 24, 39] cameras.

Shah and Aggarwal [78] designed a calibration method for fish-eye lens exploiting a calibration pattern with known 3D positions of target marks. Effective focal length, pixel size, radial and tangential distortions are estimated. The image of the optical center is determined by a laser beam. They used polynomial model of the 5^{th} degree for both kinds of distortions. The calibration is based on Lagrange minimization. Beauchemin et al. [9] and Bakstein and Pajdla [4] proposed a similar method in comparison to [78] for fish-eye lens calibration, they used a calibration pattern with known structure. In [4], a four-parametric non-polynomial model, in [9] a polynomial model capturing non-linear projection of scene to a sensor is assumed. The camera model (intrinsic and extrinsic parameters) is obtained by minimizing reprojection errors.

Devernay and Faugeras [22], Pajdla et al. [74], and Zhang [99] exploited the fact that the lines in space have to be viewed by a narrow-angle view pinhole camera as lines and that lens distortion changes the lines to curves. The camera is calibrated by correcting curves to lines, known as the plumb line method. Zhang [99] considered lens distortion as an integral part of a camera model and dealt with the epipolar geometry between two radially distorted images.

Swaminathan and Nayar [90] and Bräuer-Burchardt and Voss [12] dealt with removing of nonlinear lens distortion in images taken by super-wide angle, fish-eye lenses and polycameras. They modificated and generalized the plumb line method for such omnidirectional cameras. Geyer and Daniilidis [34, 37] proposed the plumb line method for para-catadioptric camera allowing to obtain all intrinsic parameters from the images of three lines without any metric information. Barreto and Araujo [6] enlarged the calibration from three lines to all catadioptric cameras. Ying and Hu [97] continued in

Geyer and Daniilidis' work. They extended their metric calibration using lines as well as spheres in the scene that project to image as conics to calibrate the catadioptric cameras.

We are more interested in the second group of the calibration methods, i.e., in the calibration of an omnidirectional camera without a knowledge of the scene.

Stein [81] designed a method for estimating radial distortion of perspective cameras from point correspondences by minimizing reprojection errors. He used epipolar geometry between three images and the trilinear constraint. A polynomial model of the distortion and square pixels were assumed. An iteration method lead to obtaining the image of the projection center and parameters of the distortion.

Xiong and Turkowski [96] calibrated fish-eye lenses by using four images related by a homography obtained by pure rotation of a camera. A cubic polynomial model, equidistance projection, and known image of the optical center were assumed. The final registration of the images lead to obtaining a $360^\circ \times 360^\circ$ omnidirectional camera.

Farid and Popescu [25] described a blind removal of image non-linearities in the absence of any calibration information or explicit knowledge of the imaging device. Their approach exploits the fact that a non-linearity introduces specific higher-order correlations in the frequency domain. This method is strongly dependent on the structure in scene and therefore usually unusable. Moreover, the method is designed for perspective cameras only.

Kang [54] introduced calibration of para-catadioptric cameras from epipolar geometry based on minimization of point distances to the epipolar curves without any calibration object, knowledge of camera motion, or knowledge of scene geometry. Our methods [62, 66] are philosophically the same but they provide closed-form solutions with capability to incorporate the calibration process in a 9- or 15-point RANSAC robust estimation technique. In [64, 63], we have shown how the convergence of RANSAC can be speeded up by using hierarchy of models, by excluding the correspondences near the image center and by using a bucketing technique. That enables to use automatically established point correspondences conataminated by mismatches.

Fitzgibbon [30] dealt with the problem of nonlinear lens distortion in the context of conventional camera self-calibration and structure from motion. He used a one parametric division model for radial distortion, introduced in [12], and suggested an algorithm for simultaneous estimation of two-view geometry and lens distortion from point correspondences. This model, however, cannot be directly used for omnidirectional cameras with angle of view larger than or even close to 180° because it represents images by points in which rays of a camera intersect an image plane. In our work [62, 66], we generalized Fitzgibbon's method to omnidirectional cameras, derived an appropriate omnidirectional model incorporating lens nonlinearity, and proposed an algorithm for estimating model parameters from epipolar geometry.

Geyer and Daniilidis [36, 39] proposed to calibrate para-catadioptric camera from an image of the absolute conic. It was shown that Euclidean reconstruction was feasible from two views with constant parameters and from three views with varying parameters.

Fabrizio et al. [24] proposed a method for calibrating a catadioptric sensor from mirror

boundaries. They assumed that mirror parameters were known and only intrinsic and extrinsic parameters of the CCD camera were to be recovered. The unknown parameters were estimated from the image of mirror boundaries.

We have shown in [62, 66, 65] that no information about the scene needs to be assumed to calibrate an omnidirectional (dioptric or catadioptric) camera. During the calibration process, the intrinsic camera parameters, the essential matrix, and correctly validated tentative point matches between two images are obtained. Remind, that our method works for catadioptric (shown on parabolic, hyperbolic, spherical mirror) as well as for dioptric (fish-eye lenses) cameras.

3.3. 3D reconstruction from omnidirectional images

Previous work related to 3D reconstruction from omnidirectional images assumed uncalibrated [83, 36, 39, 38, 80] or usually calibrated catadioptric sensors [53, 55, 68, 17, 23] or assumed some additional information about the motion [59, 16].

One of the earliest work by Ishiguro et al. [53] in 1992 on non-central panoramic images described an omnidirectional stereo system which uses two panoramic views to construct 3D reconstruction. Each panoramic view was created by rotating a vertical slit. Kang and Szeliski [55] followed a similar idea. They created a central omnidirectional camera by merging vertical slits of a rotated calibrated conventional camera and showed the technique for omnidirectional depth extraction.

Nene and Nayar [68] proposed the use of mirrors (planar, ellipsoidal, hyperboloidal and paraboloidal) and a single camera for computational stereo. They assumed calibrated cameras to obtain 3D metric reconstruction. The similar work has been done by Chang and Herbert [17], they showed the performance of the mirrors in SFM estimation.

Sturm [83] proposed a method for interactive 3D reconstruction of piecewise planar objects from a single panoramic view. The simple calibration of a para-catadioptric camera from a single view using known mirror parameter was indicated. The 3D reconstruction was done using geometrical constraints (coplanarity, perpendicularity and parallelism) provided by the user.

Doubek and Svoboda [23] proposed a scheme for the reliable reconstruction of indoor scenes from few catadioptric images. A set of hand-detected correspondeces was established across (not necessarily all) images. The parameters of the catadioptric sensor were approximately known and no precise self-calibration method was used. They showed that a reliable 3D reconstruction is possible even without complicated non-linear self calibation and/or reconstruction methods.

The 3D reconstruction from large sets of calibrated omnidirectional images with help of GPS was introduced by Mellor [59]. Similarly, Bunschoten and Kröse [16] described a multi-baseline stereo algorithm for 3D reconstruction of an environment from a set of calibrated panoramic images with known relative camera positions and catadioptric (hyperbolic mirror) camera parameters.

Geyer and Daniilidis [38] proposed a method for 3D conformal rectification of para-

catadioptric stereo pair in order to use it in stereo matching. The rectification is shown to be conformal in that it is locally distortionless.

Spacek [80] suggested non-central omnidirectional stereo using two coaxial conical mirrors. He found some advantages over catadatioptric cameras using curved mirrors and gave simple theory discribing such a setup. The benefit of the coaxial omnidiretional stereo is rather practical than theoretical.

In our work [60, 66, 65], we demonstrated automatic 3D metric reconstructions from two and many uncalibrated omnidirectional cameras. For many cameras we applied a projective factorization technique. In contrary to [83], where user interaction was needed, the reconstruction is computed from automatically detected point correspondences validated by an auto-calibration process. We showed in [65] that the auto-calibration method designed for *central* catadioptric omnidirectional cameras can be employed in a hierarchy from uncalibrated slightly *non-central* catadioptric images to SFM estimation using point correspondences and epipolar geometry only.

4 Geometry of Central Omnidirectional Cameras

Omnidirectional cameras have a large field of view, therefore, they have to be treated in different way as standard perspective cameras with a narrow angle of view. We show that the standard perspective model is restricted and cannot be used for omnidirectional cameras. We introduce a model allowing to represent omnidirectional cameras and thus all scene points around the camera. We describe the model capturing the formation of the digital image and propose all steps leading to a calibrated omnidirectional camera from digital images.

4.1. Omnidirectional projection

Standard perspective camera model maps all scene points $\mathbf{X}$ from a *line* passing through an optical center of a camera to *one* image point $\mathbf{x}$, see Figure 4.2(a), so that [26, 43]

$$\exists \alpha \neq 0 : \ \alpha \, \mathbf{x} = \mathtt{P} \, \mathbf{X}, \tag{4.1}$$

where $\mathtt{P} \in \mathbb{R}^{3\times 4}$ is a projection matrix, $\mathbf{X} \in \mathbb{R}^4 \setminus \{0\}$ is a scene point, and $\mathbf{x} \in \mathbb{R}^3 \setminus \{0\}$ represents an image point. The representation of image points by lines assigns to a single image point points on a ray passing through an optical center that are in front as well as behind of the camera. Thus, all points from the ray project to the same point in the projection plane. It allows to represent only scene points lying in a halfspace including an image plane as a border.

Real omnidirectional cameras with angle of view larger than 180°, however, project points in front of the camera to one point and points behind the camera to a different point, see Figure 4.1. It follows that the perspective model is sufficient for directional cameras which cannot see two halfspaces divided by the plane containing the optical center at the same time (i.e., points lying on the opposite half-lines w.r.t. the center of a camera), but cannot be used for omnidirectional cameras. The omnidirectional camera model has to take into account that every line passing through an optical center is split into two half-lines that are projected to two distinct image points (in ideal case when FOV is 360°). Therefore, omnidirectional cameras have to be represented by *half-lines*, see Figure 4.2(b).

We may represent image points, e.g., in a *spherical model*, i.e. as a set of unit vectors in $\mathbb{R}^3$ such that *one* vector corresponds just to one of half one-dimensional subspaces of $\mathbb{R}^3$. It means that *one* image point represents all scene points lying on a half-line

emanating from a camera center in contrary to the perspective model, where one image point represents all scene points lying on whole line passing through the optical center. The projection equation for omnidirectional cameras reads as

$$\exists \alpha > 0 : \ \alpha \, \mathbf{q} = \mathtt{P} \, \mathbf{X}, \tag{4.2}$$

where $\mathtt{P}$, $\mathbf{X}$ are the same as in Equation (4.1) and $\mathbf{q} \in \mathbb{R}^3 \setminus \{0\}$ is a 3D vector representing an image point.

With the spherical model, more can be done than with the standard perspective model. It is possible to obtain stronger form of epipolar constraint [94], constraint on five points in two images [92, 93], and the epipolar geometry can be augmented by an orientation [19].

4.2. Image formation

By image formation we will understand the formation of a digital image from a surrounding scene through an optics (including mirrors) and a digitization process.

In the next, we will assume that the lenses and mirrors are

i) symmetric w.r.t. an axis and

ii) the axis of the lens, or the mirror, is perpendicular to a sensor plane,

see Figure 4.3. The axial symmetry is guaranteed by manufacturing. The perpendicularity of the lens axis and the sensor plane is often guaranteed by camera construction since lenses are mounted directly on cameras. For catadioptric cameras, the perpendicularity of the sensor plane to the mirror axis is achieved by placing a perspective camera above the mirror such that the axes of the mirror and the camera are parallel. The position of a camera center w.r.t. a mirror focal point is restricted for each type of quadric mirror to obtain central projection. Let us emphasize that the perpendicularity of a sensor is not necessary to obtain central projection but allows us to recover affine transformation caused by the digitization process (will be described later).

In the following, we assume central omnidirectional cameras. A fisheye lens is used in illustrative images, however, equations hold for both dioptric and catadioptric cameras.

Suppose we are observing a scene point $\mathbf{X}$ by an omnidirectional camera, see Figure 4.4. We use the spherical model explained in Section 4.1. The projection of the scene point $\mathbf{X}$ on the unit sphere around the projection center $\mathbf{C}$ is represented by unit vectors $\mathbf{q}'' \in S^3 = \left\{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\right\}$. There is always a (possibly non-unit) vector $\mathbf{p}'' = \left(\mathbf{x}''^\top, \, z''\right)^\top$ with the same direction as $\mathbf{q}''$, which maps to the *sensor plane* point $\mathbf{u}''$ so that $\mathbf{u}''$ is *collinear* with $\mathbf{x}''$, i.e.

$$\mathbf{p}'' = \begin{pmatrix} h(\|\mathbf{u}''\|, \mathbf{a}'') \, \mathbf{u}'' \\ g(\|\mathbf{u}''\|, \mathbf{a}'') \end{pmatrix}, \tag{4.3}$$

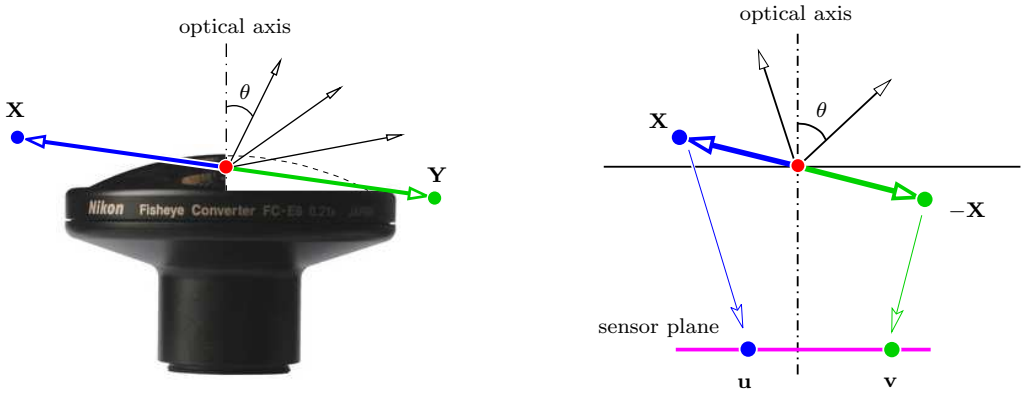
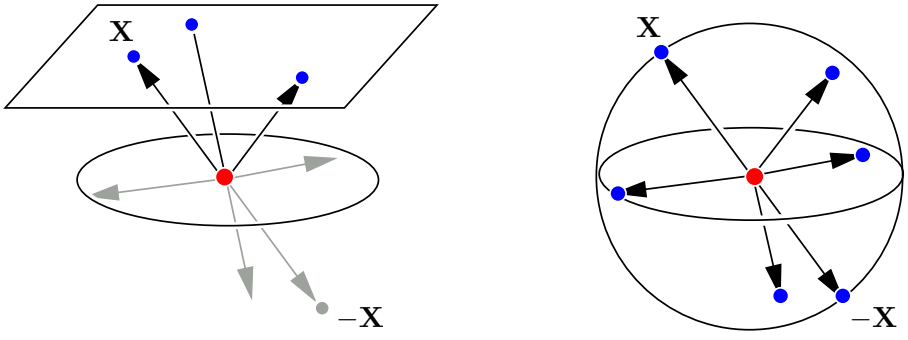


Figure 4.1.: Omnidirectional projection of *two* scene points $\mathbf{X}$, $\mathbf{Y} = -\mathbf{X}$ lying on opposite half-lines to a sensor plane as *two* different image points $\mathbf{u}$, $\mathbf{v}$.



(a) (b)

Figure 4.2.: Central camera models. (a) The standard perspective model *not* distinguishing scene points lying on opposite half-lines. (b) The spherical model distinguishing two scene points lying on opposite half-lines.

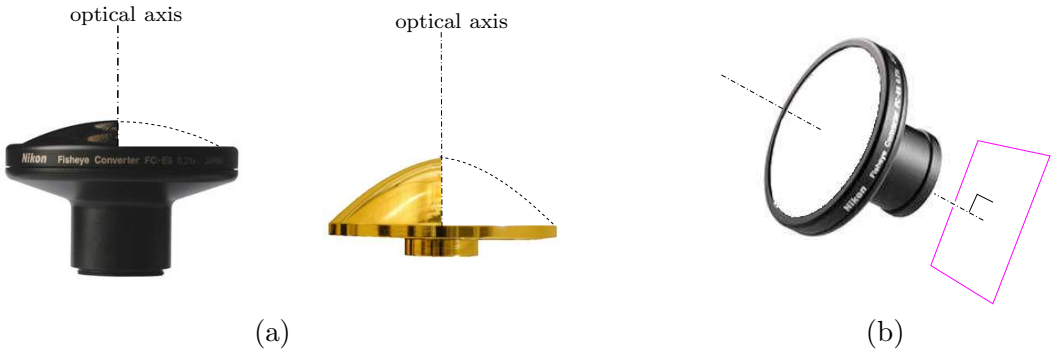


Figure 4.3.: Two assumptions in an image formation: (a) lenses and mirrors are symmetric and (b) a sensor plane is perpendicular to the axis of symmetry.

where g, h are functions $\mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$, which depend on the radius $\|\mathbf{u}''\|$ of the sensor point w.r.t. to the image of the optical axis (the center of symmetry) and on some parameters $\mathbf{a}'' \in \mathbb{R}^N$. N is the number of parameters. We will somewhere write $h(\|\mathbf{u}''\|)$ instead of $h(\|\mathbf{u}''\|, \mathbf{a}'')$ for simplicity. The collinearity of $\mathbf{u}''$ with $\mathbf{x}''$, i.e., $\mathbf{x}'' = \gamma\,\mathbf{u}''$ is clear from Figure 4.4(b), where $\gamma = h(\|\mathbf{u}''\|)$ and thus $\mathbf{x}'' = h(\|\mathbf{u}''\|)\,\mathbf{u}''$.

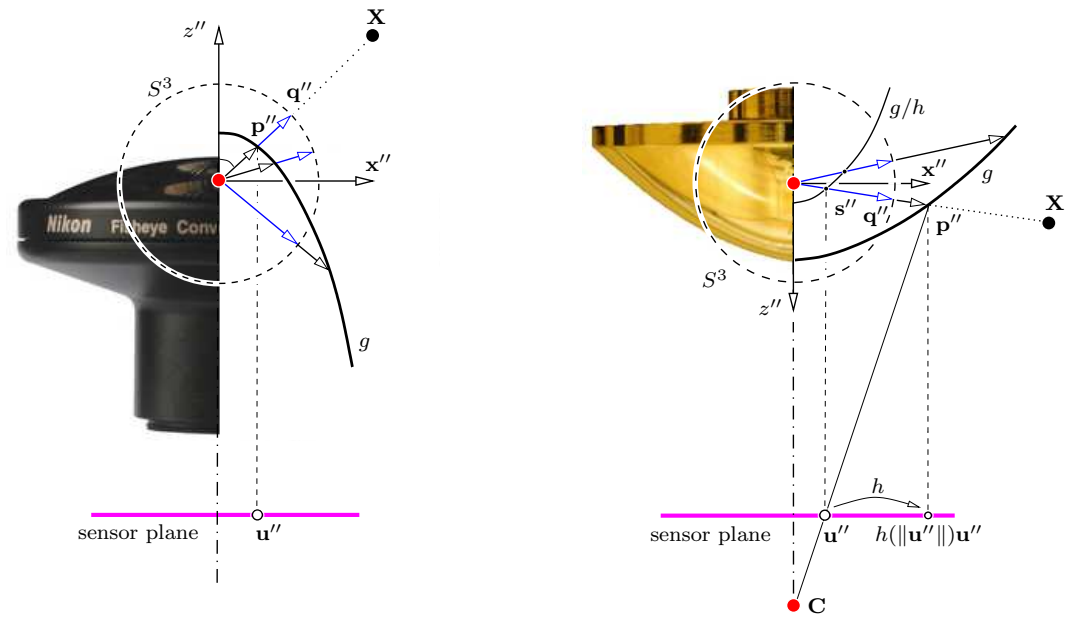
The functions g, h are *rotationally symmetric* due to the assumptions made at the beginning of this Section (lens, mirror symmetry and sensor plane perpendicularity) and thus $g, h(\|\mathbf{u}''\|) = g, h(\|\mathtt{R}\,\mathbf{u}''\|)$ holds for every rotation $\mathtt{R} \in \mathbb{R}^{2\times 2}$ in the sensor plane around the center of symmetry.

The functions g, h differ for various types of lenses and mirrors. For lenses, the functions depend on the type of the lens projection (equisolid, equiangular, etc.) and for the mirrors they depend on the shape of the mirror (parabolic, hyperbolic, elliptical). The mapping of the vector $\mathbf{p}''$ to the sensor plane point $\mathbf{u}''$ through the functions g, h is shown in Figure 4.4 for a fish-eye lens and a hyperbolic mirror. There holds for all fish-eye lenses that the function $h = 1$ and the vector $\mathbf{p}''$ is mapped orthographically to the sensor plane.

In the case of mirrors, the vector $\mathbf{p}''$ is mapped by a perspective camera with the optical center $\mathbf{C}$ (which can lie at infinity) to the point $\mathbf{u}''$. We can also say that the vector $\mathbf{p}''$ is projected orthographically to the sensor plane to the point $h(\|\mathbf{u}''\|)\,\mathbf{u}''$ or that $\mathbf{p}''$ is represented by $\mathbf{s}''$ (obtained as the intersection of function g/h and $\mathbf{p}''$) is projected orthographically to the sensor plane point $\mathbf{u}''$.

At this moment, the following question may arise: Why do we need both functions g, h and not just one g/h? The function g may have physical meaning, e.g., it can capture the shape of a mirror. The function h represents the projection of the camera, e.g., for the orthographic projection $h = 1$. However, for some shapes of mirrors, there can arise a situation where $h(\|\mathbf{u}''\|) = 0$ for $\|\mathbf{u}\| \neq 0$ and the division g/h can not be calculated. It means that the model would give the vector $\mathbf{p}'' = (0,\, 0,\, 1)^\top$ for $\|\mathbf{u}''\| \neq 0$.

The mapping can be specialized to obtain the standard perspective projection and



(a) (b)

Figure 4.4.: The mapping of a scene point $\mathbf{X}$ into a sensor plane to a point $\mathbf{u}''$ for (a) a fish-eye lens and (b) a hyperbolic mirror.

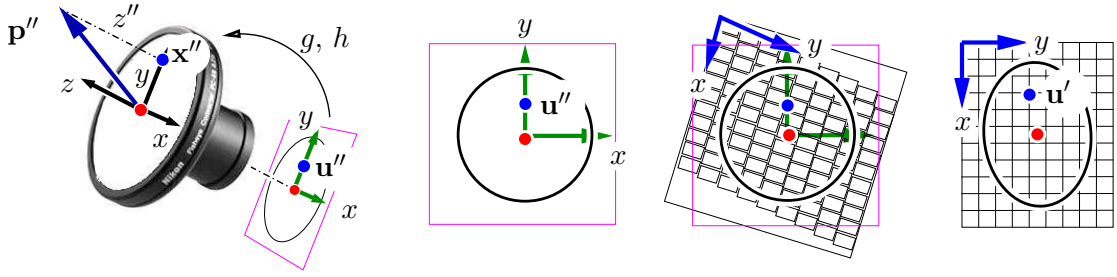


Figure 4.5.: Omnidirectional image formation. (a) A lens projects scene points to a sensor plane. (b) The sensor plane with view field circle. (c) Digitization process. (Courtesy of H. Bakstein) (d) An acquired image related to the image on the sensor plane by an affine transformation. The view field circle is transformed to an ellipse.

omnidirectional projection as follows

$$\begin{aligned} &\text{perspective projection:} && \begin{pmatrix} 1\,\mathbf{u}'' \\ 1 \end{pmatrix} \\ &\text{omnidirectional projection:} && \begin{pmatrix} h(\|\mathbf{u}''\|)\,\mathbf{u}'' \\ g(\|\mathbf{u}''\|) \end{pmatrix}. \end{aligned} \tag{4.4}$$

It was shown and proved in [35] that every cata-dioptric (parabolic, hyperbolic, elliptical), stereographic and standard perspective projection is equivalent to the projection onto the unit sphere followed by a projection the sphere to a plane with a projection center on the perpendicular to the plane. Every such mapping can be re-written to the form of Equation (4.4) with

$$\begin{aligned} h(\|\mathbf{u}''\|) &= \frac{l(l+m)+\sqrt{\|\mathbf{u}''\|^2(1-l^2)+(l+m)^2}}{\|\mathbf{u}''\|^2+(l+m)^2}, \\ g(\|\mathbf{u}''\|) &= \frac{l\|\mathbf{u}''\|^2+(l+m)\sqrt{\|\mathbf{u}''\|^2(1-l^2)+(l+m)^2}}{\|\mathbf{u}''\|^2+(l+m)^2}, \end{aligned}$$

where constants l, m, depending on the type of projection, can be found in [35]. The Equation (4.4) can be thus regarded as a suitable representation of central omnidirectional projections covering catadioptric and dioptric cameras.

Until now we were describing projections of a scene onto a sensor plane, see Figure 4.5(a, b). To acquire an image from the sensor plane, some digitization process has to be done. Suppose a general digital sensor composed of non-square pixels aligned in a "linear" but non-rectangular grid (non zero skew in the standard terminology of [43]), see Figure 4.5(c). The digitization process transforms, see Figure 4.5(d), the circular field

of view into an elliptical one in the digital image. The digitization can be represented by an affine transformation

$$\mathbf{u}'' = \mathtt{A}'\mathbf{u}' + \mathbf{t}', \tag{4.5}$$

where $\mathbf{u}'$ is a point in a digital image, $\mathtt{A}' \in \mathbb{R}^{2\times 2}$ is a regular matrix and $\mathbf{t}' \in \mathbb{R}^2$ is a translational vector.

The complete image formation model capturing the projection of a scene point $\mathbf{X}$ into the digital image point $\mathbf{u}'$ can be divided into three parts:

i) a central projection of the scene point $\mathbf{X}$ to the vector $\mathbf{p}''$,

ii) a non-perspective optics or mirror reflection, described by the functions g and h, mapping $\mathbf{p}''$ to $\mathbf{u}''$, and

iii) a digitization process transforming the sensor plane point $\mathbf{u}''$ to the digital image point $\mathbf{u}'$.

The complete image formation can be written as

$$\frac{1}{\alpha''}\,\mathtt{P}''\,\mathbf{X} = \mathbf{p}'' = \begin{pmatrix} \mathbf{x}'' \\ z'' \end{pmatrix} = \begin{pmatrix} h(\|\mathbf{u}''\|)\mathbf{u}'' \\ g(\|\mathbf{u}''\|) \end{pmatrix} = \begin{pmatrix} h(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|)(\mathtt{A}'\mathbf{u}' + \mathbf{t}') \\ g(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|) \end{pmatrix},$$

so that the projection equation for omnidirectional cameras is

$$\exists\, \alpha'' > 0\colon \quad \alpha'' \begin{pmatrix} h(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|)(\mathtt{A}'\mathbf{u}' + \mathbf{t}') \\ g(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|) \end{pmatrix} = \mathtt{P}''\,\mathbf{X}\,, \tag{4.6}$$

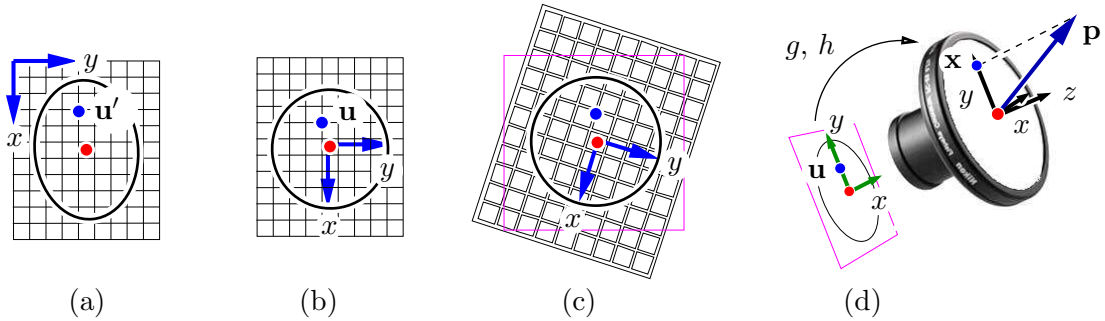
where $\mathtt{P}'' \in \mathbb{R}^{3\times 4}$ is a projection matrix, $\mathtt{A}' \in \mathbb{R}^{2\times 2}$, $\operatorname{rank}(\mathtt{A}') = 2$, and $\mathbf{t}' \in \mathbb{R}^2$ represent an affine transformation in the sensor plane and $\mathbf{u}' \in \mathbb{R}^2$ is a point in the digital image.

4.3. Camera Calibration

The aim of the calibration is to find a mapping from a digital image point $\mathbf{u}'$ to a corresponding 3D ray $\mathbf{p}''$. Normalizing $\mathbf{p}''$ gives $\mathbf{q}''$ with unit length. The calibration can be divided into two steps. First, the mapping from the digital image to the sensor plane caused by a digitization has to be recovered. Second, the mapping from a sensor plane to scene rays caused by optics (or mirror reflection) has to be found.

The advantage of omnidirectional cameras over the standard directional cameras is that the full view field circle is projected into the sensor plane and thus observable in the digital image as the view field ellipse, see Figure 4.6(a). Usually, the ellipse is very close to a circle since the camera skew parameter is negligible and pixels are nearly square.

The point $\mathbf{u}'$ in an acquired digital image is related to the point in the sensor plane $\mathbf{u}''$ by Equation (4.5), what corresponds to the mapping of a circle to an ellipse in an image formation step.



(a) (b) (c) (d)

Figure 4.6.: Omnidirectional camera calibration process. (a) An acquired digital image with an elliptical field of view. (b) A transformed image with a circular field of view. (c) Simulated sensor with zero skew parameter and square pixels. (d) Projection of a sensor plane to 3D rays.

We can transform the view field ellipse, which can often be easily obtained by fitting an ellipse to the boundary of the view field in the digital image, to a circle. The transformation can be written again as

$$\mathbf{u} = \mathtt{A}\,\mathbf{u}' + \mathbf{t}\,, \tag{4.7}$$

where $\mathtt{A} \in \mathbb{R}^{2\times 2}$ and $\mathbf{t} \in \mathbb{R}^2$.

However, there is an ambiguity in the affine transformation with two degrees of freedom left by mapping an allipse to a circle. We do not know the radius of the circle and the rotation $\mathtt{R} \in \mathbb{R}^{2\times 2}$ around the circle center is not fixed. Denote the change of the radius by a scale factor $\rho > 0$. The scale factor ρ captures the relationship between camera Cartesian coordinate system (measured in pixels) and sensor Cartesian coordinate system (measured ,e.g., in millimetres). To summarize the relationship between $\mathtt{A}', \mathbf{t}'$, and $\mathtt{A}, \mathbf{t}$ the following holds

$$\begin{aligned} \mathbf{t} &= \frac{1}{\rho}\,\mathtt{R}^{-1}\,\mathbf{t}'\,, \\ \mathtt{A} &= \frac{1}{\rho}\,\mathtt{R}^{-1}\,\mathtt{A}'. \end{aligned} \tag{4.8}$$

The matrix $\mathtt{R}$ and the scale ρ are unknown since they cannot be recovered from the view field ellipse.

The process of determining the affine matrix $\mathtt{A}$ and the vector $\mathbf{t}$ is called the *pre-calibration* step in the rest of the thesis. The pre-calibrated image corresponds to the image acquired by a camera with zero skew parameter and square pixels, see Figure 4.6(c). After the pre-calibration step, a radially symmetric image is obtained. It means that in such a pre-calibrated image a non-linearity (mapping an image point to a 3D ray) can be modelled in one direction from the origin to the border of the image.

Equation (4.8) gives the relationship between the point $\mathbf{u}''$ in the sensor plane and the point $\mathbf{u}$ in the pre-calibrated image, i.e.,

$$\begin{array}{c} \mathbf{u}'' = \mathtt{A}'\,\mathbf{u}' + \mathbf{t}' \\ \mathbf{u} = \mathtt{A}\,\mathbf{u}' + \mathbf{t} \end{array} \quad \Rightarrow \quad \mathbf{u}'' = \rho\,\mathtt{R}\,\mathbf{u}. \tag{4.9}$$

It means that we are able to compute the point $\mathbf{u}$ in a camera Cartesian coordinate system up to a scale and a rotation w.r.t. the point $\mathbf{u}''$ in a sensor Cartesian coordinate system.

The second calibration step is to compute a corresponding 3D ray $\mathbf{p}$ for a point $\mathbf{u}$ in the pre-calibrated image and to find the relationship between the vector $\mathbf{p}$ and $\mathbf{p}''$. We modelled the mapping $\mathbf{u}'' \xrightarrow{g,h} \mathbf{p}''$ in Equation (4.3) but now we need the mapping $\mathbf{u} \to \mathbf{p}$. It follows from Equations (4.3) and (4.9) that

$$\mathbf{p}'' = \begin{pmatrix} h(\|\mathbf{u}''\|, \mathbf{a}'')\,\mathbf{u}'' \\ g(\|\mathbf{u}''\|, \mathbf{a}'') \end{pmatrix} = \begin{pmatrix} h(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|, \mathbf{a}'')\,(\mathtt{A}'\mathbf{u}' + \mathbf{t}') \\ g(\|\mathtt{A}'\mathbf{u}' + \mathbf{t}'\|, \mathbf{a}'') \end{pmatrix} = \begin{pmatrix} h(\|\rho\mathtt{R}\mathbf{u}\|, \mathbf{a}'')\,(\rho\mathtt{R}\mathbf{u}) \\ g(\|\rho\mathtt{R}\mathbf{u}\|, \mathbf{a}'') \end{pmatrix}. \tag{4.10}$$

From the radial symmetry of the functions g, h it follows

$$\mathbf{p}'' = \begin{pmatrix} h(\|\rho\mathbf{u}\|, \mathbf{a}'')\,(\rho\mathtt{R}\mathbf{u}) \\ g(\|\rho\mathbf{u}\|, \mathbf{a}'') \end{pmatrix} = \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} h(\|\rho\mathbf{u}\|, \mathbf{a}'')\,(\rho\mathbf{u}) \\ g(\|\rho\mathbf{u}\|, \mathbf{a}'') \end{pmatrix}. \tag{4.11}$$

We cannot determine the functions $g(\|\rho\,\mathbf{u}\|, \mathbf{a}'')$, $h(\|\rho\,\mathbf{u}\|, \mathbf{a}'')$ in $\|\rho\,\mathbf{u}\|$ because we have assumed that the scale ρ is unknown. Let further assume that

$$\begin{aligned} \exists m\colon \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N, \exists i, j\colon \mathbb{R} \to \mathbb{R}, \forall \mathbf{u} \in \mathbb{R}^2, \forall \rho \in \mathbb{R}, \forall \mathbf{a}'' \in \mathbb{R}^N\colon \\ h(\|\rho\mathbf{u}\|, \mathbf{a}'') = i(\rho) h\Big(\|\mathbf{u}\|, m(\rho, \mathbf{a}'')\Big), \\ g(\|\rho\mathbf{u}\|, \mathbf{a}'') = j(\rho) g\Big(\|\mathbf{u}\|, m(\rho, \mathbf{a}'')\Big), \end{aligned} \tag{4.12}$$

where, let us remind it, N is the number of parameters. Denote $\mathbf{a} = m(\rho, \mathbf{a}'')$. Then the following can be written

$$\begin{aligned} h(\|\rho\mathbf{u}\|, \mathbf{a}'') &= i(\rho) h(\|\mathbf{u}\|, \mathbf{a}), \\ g(\|\rho\mathbf{u}\|, \mathbf{a}'') &= j(\rho) g(\|\mathbf{u}\|, \mathbf{a}). \end{aligned}$$

Equation (4.11) becomes

$$\mathbf{p}'' = \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \rho\, i(\rho)\, h(\|\mathbf{u}\|, \mathbf{a})\,\mathbf{u} \\ j(\rho)\, g(\|\mathbf{u}\|, \mathbf{a}) \end{pmatrix}. \tag{4.13}$$

Assume, again, that it holds

$$\forall \rho \in \mathbb{R}\colon\ \rho\, i(\rho) = j(\rho), \tag{4.14}$$

then Equation (4.13) becomes

$$\mathbf{p}'' = j(\rho) \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} h(\|\mathbf{u}\|, \mathbf{a})\, \mathbf{u} \\ g(\|\mathbf{u}\|, \mathbf{a}) \end{pmatrix}.$$

Finally we end up with the relationship between the vector $\mathbf{p}''$ in the sensor Cartesian coordinate system and the calibrated vector $\mathbf{p}$ in the camera Cartesian coordinate system

$$\mathbf{p}'' \simeq \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}\,. \tag{4.15}$$

The important result is that the functions g, h, mapping the point $\mathbf{u}''$ in the sensor plane to vector $\mathbf{p}''$ through parameters $\mathbf{a}''$, i.e., $\mathbf{u}'' \xrightarrow{g,h(\|\mathbf{u}''\|,\mathbf{a}'')} \mathbf{p}''$, can be used for mapping the point $\mathbf{u}$ in the pre-calibrated image to the vector $\mathbf{p}$, i.e. $\mathbf{u} \xrightarrow{g,h(\|\mathbf{u}\|,\mathbf{a})} \mathbf{p}$, only the values of parameters are changed. The vectors $\mathbf{p}''$ and $\mathbf{p}$ differ in length and they are mutually rotated around the optical axis. The ambiguity in the length and the rotation does not affect angles between the vectors. Therefore, a *metrically* calibrated camera is obtained.

Assuming "only" $\forall \rho \in \mathbb{R}\colon\ i(\rho) = j(\rho)$ instead of Equation (4.14) would lead to a specific non-metric calibration and subsequently to a slightly non-metric 3D reconstruction. It is the same as unknown focal length in a standard camera calibration.

An important theorem about omnidirectional camera calibration can be formulated.

Theorem 1 *Let the following holds*

$\exists m\colon \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N, \exists i\colon \mathbb{R} \to \mathbb{R}, \forall \mathbf{u} \in \mathbb{R}^2, \forall \rho \in \mathbb{R}, \forall \mathbf{a}'' \in \mathbb{R}^N\colon$

$$\begin{aligned} h(\|\rho\mathbf{u}\|, \mathbf{a}'') &= i(\rho) h\Big(\|\mathbf{u}\|, m(\rho, \mathbf{a}'')\Big)\,, \\ g(\|\rho\mathbf{u}\|, \mathbf{a}'') &= \rho\, i(\rho) f\Big(\|\mathbf{u}\|, m(\rho, \mathbf{a}'')\Big)\,, \end{aligned} \tag{4.16}$$

and let the functions g, h model a concrete lens or a mirror. Then, the change of camera resolution causes that just the values (not the number) of parameters are changed. Moreover,

$$\mathbf{p}'' \simeq \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}\,. \tag{4.17}$$

Proof. See the derivation of Equation (4.15) from Equation (4.10) above. □

The parametric forms of the functions g, h have to be known before the calibration. The calibration method will estimate the parameters $\mathbf{a}$. The following consequence of Theorem 1 can be formulated.

Consequence 1 *We are able to design a calibration method for omnidirectional cameras consisting of mirrors or lenses independently on the resolution of a sensor chip. The change of the resolution of the chip does not change the form of functions g, h.*

(a) (b) (c) (d)

Figure 4.7.: The calibration process shown on a real image. (a) An image on a sensor plane. (b) An acquired digital image with an ellipse fitted on a view field. (c) A transformed circular pre-calibrated image. (d) Representation of an image on a sphere.

We assumed the existence of functions $m(.)$ and $i(.)$ in Theorem 1. The function $m(.)$ absorbs the scale ρ into a new vector $\mathbf{a}$ with the same number of elements as $\mathbf{a}''$. The function $i(.)$ absorbs the scale ρ which arised by taking it out from functions g a h.

Consequence 2 *Suppose that we have functions g and h characterizing an omnidirectional camera. Finding functions m, i from Theorem 1 for the camera gives Consequence 1, i.e., that the calibration method for the omnidirectional camera can be designed on the resolution of the camera chip independently.*

In every section devoted to a concrete omnidirectional camera, we will show the corresponding functions g and h and we will find functions m and i.

The full calibration process is illustrated on a real image in Figure 4.7. The figure shows the digital image as a rotated ellipse since we wanted to demonstrate the effect of the digitization process. Usually, the rotation is very small and the ellipse is close to a circle. Finally, as a result of the calibration, which will be described later, the corresponding 3D vectors are obtained for every image point $\mathbf{u}'$.

4.4. Epipolar geometry

The epipolar geometry describes the geometrical relationship between a pair of central cameras observing the same rigid scene. The epipolar geometry for standard perspective directional cameras is well understood [43, 26]. Here, we review it very briefly.

Suppose a stereo-pair, consisting of standard perspective cameras, which is observing some rigid scene, see Figure 4.8(a). Scene points, together with both camera centers $\mathbf{C}_1$, $\mathbf{C}_2$, create a pencil of planes through the baseline $\overline{\mathbf{C}_1\mathbf{C}_2}$. Such planes intersect the image planes of the cameras in straight lines, called *the epipolar lines*. All epipolar lines

intersect at one point, called *the epipole*. If the epipole lies at infinity then the epipolar lines become parallel. Figure 4.8(a) shows a scene point $\mathbf{X}$ that is projected by the cameras to image points $\mathbf{u}_1$, $\mathbf{u}_2$. The epipole $\mathbf{e}_1$, resp. $\mathbf{e}_2$, represents the image of the optical center $\mathbf{C}_2$ of the second camera, resp. of the optical center $\mathbf{C}_1$ of the first one. The epipolar constraint for a pair of directional images reads as

$$\begin{pmatrix} \mathbf{u}_2^\top & 1 \end{pmatrix} \mathtt{F} \begin{pmatrix} \mathbf{u}_1 \\ 1 \end{pmatrix} = 0\,, \tag{4.18}$$

where $\mathtt{F} \in \mathbb{R}^{3\times 3}$ is a fundamental matrix with $\mathrm{rank}(\mathtt{F}) = 2$. The fundamental matrix represents a mapping which maps a point (point in $\mathbb{P}^2$, 1D subspace in $\mathbb{R}^3$) in the first image plane to a line (line in $\mathbb{P}^2$, 2D subspace in $\mathbb{R}^3$) in the second image plane.

The epipolar geometry can be formulated for central omnidirectional central cameras, i.e., for catadioptric [86, 73, 85] and for dioptric (with fish-eye lenses) omnidirectional cameras.

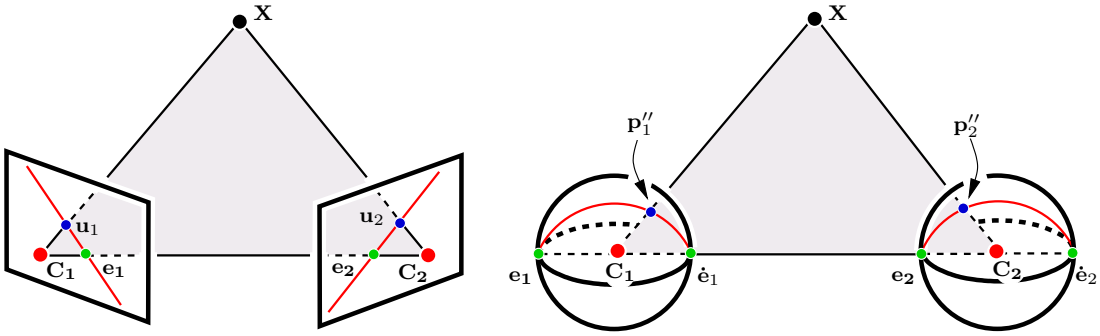
The difference between directional and omnidirectional cameras, which has impact on epipolar geometry, lies in distinguishability of ray orientation, Section 4.1. In the case of the perspective cameras, epipolar planes intersect a planar image retina in *lines* crossing only *one* epipole, see Figure 4.8(a). In the case of omnidirectional cameras, the epipolar planes intersect a spherical retina in circles, which are projected to the sensor plane as *curves* crossing *two epipoles*, see Figure 4.8(b). The curves are conics for quadric catadioptric cameras [85], but in general, e.g., for fish-eye lenses they are more general curves. There are two epipoles, hence omnidirectional cameras distinguish between lines and half-lines, see positive α in Equation (4.2), and project the baseline $\overline{\mathbf{C}_1\mathbf{C}_2}$ as two different image points, see Section 4.1. It depends on the camera field of view, how many epipoles are seen in the image. The number of epipoles varies from zero to two, see Figure 4.9.

Figure 4.8(b) shows how the epipolar plane containing a scene point $\mathbf{X}$ that intersects spherical retinas, which represent the omnidirectional cameras. There are two points of intersection, the epipoles $\mathbf{e}_1$, $\dot{\mathbf{e}}_1$ in the first camera, resp. the epipoles $\mathbf{e}_2$, $\dot{\mathbf{e}}_2$ in the second camera. The vectors $\mathbf{p}_1''$, $\mathbf{p}_2''$ represent the scene point $\mathbf{X}$ in the first and second camera, respectively. Figure 4.9 shows how the circles are projected to a digital image. Notice, that for every vector $\mathbf{p}_1$, projected into an image as $\mathbf{u}_1$, there is a curve in the second image on which the corresponding point $\mathbf{u}_2$ lies.

The epipolar geometry can be formulated for vectors $\mathbf{p}_1''$ and $\mathbf{p}_2''$ since they create an epipolar plane. The epipolar constraint for a pair of omnidirectional images reads as

$$\mathbf{p}_2''^\top \mathtt{F}'' \mathbf{p}_1'' = 0\,, \tag{4.19}$$

where $\mathtt{F}''$ has the same meaning as in Equation (4.18) and represents the same mapping, i.e. the mapping of a one-dimensional subspace to a two-dimensional subspace in $\mathbb{R}^3$. Equation (4.19) shows how the epipolar geometry can be employed for omnidirectional cameras. Notice that the epipolar constraint cannot be applied directly to image points as it is in the case of standard perspective directional cameras, but has to be applied to 3D vectors computed from image points using functions g, h.



(a) (b)

Figure 4.8.: Epipolar geometry for central cameras. (a) Perspective cameras with a planar retina. (b) The omnidirectional central cameras represented by spherical retinas.

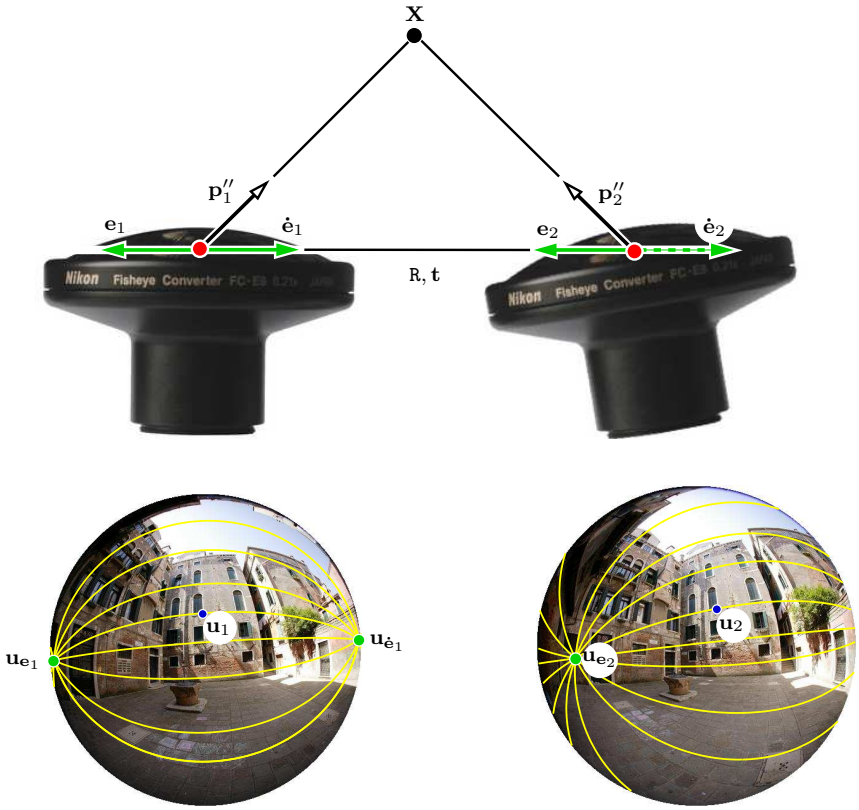


Figure 4.9.: Epipolar geometry for a pair of omnidirectional cameras consisting of a fish-eye lens. Epipolar curves are depicted in the pre-calibrated images as yellow curves.

We have shown in Section 4.3 that we are able to obtain vectors $\mathbf{p}_1$, $\mathbf{p}_2$ that are related to $\mathbf{p}''_1$, $\mathbf{p}''_2$ by Equation (4.15). Substituting Equation (4.15) into Equation (4.19) leads to the equivalent constraint

$$\mathbf{p}_2^\top \underbrace{\begin{pmatrix} \mathtt{R}_2 & \\ & 1 \end{pmatrix}^\top \mathtt{F}'' \begin{pmatrix} \mathtt{R}_1 & \\ & 1 \end{pmatrix}}_{\mathtt{F}} \mathbf{p}_1 = 0$$

$$\mathbf{p}_2^\top \mathtt{F}\, \mathbf{p}_1 = 0\,. \tag{4.20}$$

The following theorem can be formulated from Equation (4.20).

Theorem 2 *The epipolar constraint holds for the vectors $\mathbf{p}_1$, $\mathbf{p}_2$ obtained from a calibration process if and only if it holds for the vectors $\mathbf{p}''_1$, $\mathbf{p}''_2$.*

As we explained in Section 4.3, the knowledge of the functions g, h and their parameters $\mathbf{a}$ allows to compute 3D vectors $\mathbf{p}$ up to a scale and a rotation w.r.t. vectors $\mathbf{p}''$. The parameters can be obtained, e.g., by some calibration method. We designed a robust auto-calibration method from point correspondences and epipolar geometry applicable for a large class of central omnidirectional cameras, described and discussed later.

We will focus on the estimation of epipolar geometry from point correspondences based on a RANSAC strategy to eliminate the incorrect correspondences. Therefore the most important component of such an algorithm is the test which is used to mark each match as inlier or outlier.

Since there is noise in correspondences (assume Gaussian), the corresponding 3D rays $\mathbf{p}_1$ in the first image and $\mathbf{p}_2$ in the second one are not exactly coplanar. The matrix $\mathtt{F}$ can be computed from some $\mathbf{p}_1$, $\mathbf{p}_2$ using 7- or 8- point algorithm [43] to obtain a linear estimate of $\mathtt{F}$. However, the number $\mathbf{p}_2^\top \mathtt{F}\mathbf{p}_1$ (algebraic error) does not equal zero for all 3D rays. The optimal measure of error of a ray w.r.t. the estimated $\mathtt{F}$ (geometric error) is given by

$$\epsilon(\mathbf{u}_1, \mathbf{u}_2, \mathtt{F}, \mathbf{a}) = \min_{\{\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2 | \hat{\mathbf{p}}_2^\top \mathtt{F} \hat{\mathbf{p}}_1 = 0\}} \|\hat{\mathbf{u}}_2 - \mathbf{u}_2\|^2 + \|\hat{\mathbf{u}}_1 - \mathbf{u}_1\|^2\,, \tag{4.21}$$

where $\mathbf{u}_1$, $\mathbf{u}_2$ are image points, $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$ are corrected image points such that epipolar geometry with fundamental matrix $\mathtt{F}$ holds for vectors $\hat{\mathbf{p}}_1$, $\hat{\mathbf{p}}_2$ computed from $\hat{\mathbf{u}}_1$, $\hat{\mathbf{u}}_2$ using parameters $\mathbf{a}$. There is a direct solution of the error in Equation (4.21), given by Hartley and Sturm [44], for standard perspective cameras. For omnidirectional cameras there is no direct solution (at least unknown in the community) since there are non-linear functions g, h to compute $\mathbf{p}$'s from $\mathbf{u}$'s. Therefore, the computation of this error is not very practical since we would have to project 3D vectors back to the image plane and minimize distances from epipolar curves. We are more interested in some error for which the direct solution is known.

If we know parameters $\mathbf{a}$ of functions g, h or we estimate the parameters together with epipolar geometry (our case, discussed in the next Chapter) then we recover correct angles between rays $\mathbf{p}$ and the optical axis. It allows us to use an angle-based measure

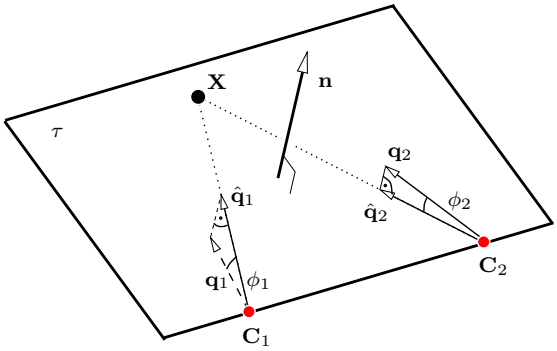


Figure 4.10.: Angular error shown on one epipolar plane τ. $\mathbf{X}$ denotes a 3D scene point, $\mathbf{n}$ is the unit normal vector to the epipolar plane, $\mathbf{C}_1$, $\mathbf{C}_2$ are camera centers, $\mathbf{q}_1$, $\mathbf{q}_2$ are rays corresponding to image points, $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_2$ are their optimal estimates lying in the epipolar plane, ϕ_1, ϕ_2 are angles between image points and their optimal estimates.

of the error with a closed-form solution. The matrix $\mathtt{F}$ in Equation (4.20) becomes an essential matrix with equal singular values [43].

If the essential matrix is estimated then the angular error given by Oliensis [70] can be used. It is argued in [70] that angular error gives for omnidirectional cameras better results compared to the error in Equation (4.21). The angular error is defined as the minimum (over the normal of an epipolar plane) of the sum of squared sinus of angles ϕ_1 and ϕ_2 between rays and the epipolar plane, see Figure 4.10, i.e.,

$$\begin{aligned} \epsilon(\mathbf{q}_1, \mathbf{q}_2, \mathtt{F}) &= \min_{\mathbf{n}} \left(\sin^2 \phi_1 + \sin^2 \phi_2 \right) \\ &= \min_{\mathbf{n}} \left(|\mathbf{n}.\mathbf{q}_1|^2 + |\mathbf{n}.\mathbf{q}_2|^2 \right) . \end{aligned} \tag{4.22}$$

Recall that $\mathbf{q}$ stands for normalized $\mathbf{p}$ of unit length. Direct solution for this error is

$$\epsilon(\mathbf{q}_1, \mathbf{q}_2, \mathtt{F}) = \frac{A}{2} - \sqrt{\frac{A^2}{4} - B} \,,$$

where

$$A = \mathbf{q}_1^\top \mathtt{F}^\top \mathtt{F} \mathbf{q}_1 + \mathbf{q}_2^\top \mathtt{F}\, \mathtt{F}^\top \mathbf{q}_2 \,, \qquad B = \left(\mathbf{q}_2^\top \mathtt{F} \mathbf{q}_1 \right)^2 .$$

Let us remind again that Oliensis' angular error can be used only if the matrix $\mathtt{F}$ is an essential matrix.

To obtain optimal estimate $\hat{\mathtt{F}}$ of the fundamental matrix, the Maximum Likelihood estimate, known as the Gold Standard method [43], using the initial estimate $\mathtt{F}$ should be applied. However, in our experiment, we use bundle adjustment to tune final 3D reconstruction what is equivalent to the Gold Standard method.

4.5. 3D reconstruction

In this section, we recapitulate a simple linear triangulation method, which generalizes the standard direct linear method [43], known as the triangulation method, to omnidirectional cameras. Usually, the image point correspondences do not exactly satisfy the geometric relations, therefore the result of the triangulation method should be regarded for an initial estimate of the scene point.

Remind from Equation (4.6) that

$$\alpha'' \mathbf{p}'' = \mathtt{P}'' \mathbf{X} \,.$$

Using the relationship between calibrated vector $\mathbf{p}$ and the "real" vector $\mathbf{p}''$ from Equation (4.15) leads to

$$\alpha \, \mathbf{p} = \underbrace{\begin{pmatrix} \mathtt{R}^\top & \\ & 1 \end{pmatrix} \mathtt{P}''}_{\mathtt{P}} \mathbf{X} = \mathtt{P} \, \mathbf{X} \,. \tag{4.23}$$

Equation (4.23) shows that unknown rotation $\mathtt{R}$ creates together with the projection matrix $\mathtt{P}''$ a new projection matrix $\mathtt{P}$. It follows that calibration ambiguity in rotation and scale causes that the scene can be reconstructed up to a scale, i.e. *metric*[1] reconstruction is obtained.

Suppose that we have calibrated cameras, i.e. we can compute 3D vectors $\mathbf{p}$ for all image points $\mathbf{u}$ such that it holds in both images

$$\begin{aligned} \alpha_1 \mathbf{p}_1 &= \mathtt{P}_1 \mathbf{X} \,, \\ \alpha_2 \mathbf{p}_2 &= \mathtt{P}_2 \mathbf{X} \,, \end{aligned} \tag{4.24}$$

where we suppose that matrices $\mathtt{P}_i$ are known. We can write the following for the first image

$$\begin{aligned} \alpha_1 x_1 &= \mathbf{r}_1^{1\top} \mathbf{X} \,, \\ \alpha_1 y_1 &= \mathbf{r}_1^{2\top} \mathbf{X} \,, \\ \alpha_1 z_1 &= \mathbf{r}_1^{3\top} \mathbf{X} \,, \end{aligned}$$

where $\mathbf{r}_1^{i\top}$ are the rows of the $\mathtt{P}_1$ and $\mathbf{p}_1 = (x_1, y_1, z_1)^\top$. The division of the equations mutually eliminates the scale α_1. The same can be done for the second image what leads to another three equations. The six equations can be combined into a form $\mathtt{A}\mathbf{X} = \mathbf{0}$,

[1] Sometimes called similarity or Euclidean.

which is an equation linear in $\mathbf{X}$. The matrix $\mathtt{A}$ is composed of

$$\mathtt{A} = \begin{bmatrix} x_1\mathbf{r}_1^{3\top} - z_1\mathbf{r}_1^{1\top} \\ x_1\mathbf{r}_1^{2\top} - y_1\mathbf{r}_1^{1\top} \\ y_1\mathbf{r}_1^{3\top} - z_1\mathbf{r}_1^{2\top} \\ x_2\mathbf{r}_2^{3\top} - z_2\mathbf{r}_2^{1\top} \\ x_2\mathbf{r}_2^{2\top} - y_2\mathbf{r}_2^{1\top} \\ y_2\mathbf{r}_2^{3\top} - z_2\mathbf{r}_2^{2\top} \end{bmatrix}, \tag{4.25}$$

where three equations have been included from each image, giving six equations in total. The matrix $\mathtt{A}$ has rank 3 (noiseless correspondences), but has six rows, instead of four as in [43]. In [43], the method is designed for vectors with the last coordinate equal to one. However, the omnidirectional cameras may have the third coordinate $z = 0$. Adding two extra rows to $\mathtt{A}$ treats this situation.

If noise is present then $\text{rank}(\mathtt{A}) \neq 3$, $\mathtt{A}\mathbf{X} = \mathbf{0}$ can be standardly solved in the least-squares sense by SVD [41]. The estimate $\mathbf{X}$ should be further used as a starting point in a nonlinear bundle adjustment minimizing reprojection errors.

4.6. Closure

In this chapter, we proposed the projection model for omnidirectional cameras. We designed appropriate representation of omnidirectional cameras and showed the difference to the standard representation used for perspective cameras. We pointed out that the standard camera model cannot be used since it does not capture view field larger that 180°.

We wrote down equations of image formation and backprojection from image point to 3D rays going from the optical center towards scene points. We showed that our omnidirectional cameras can be calibrated up to ambiguity in a scale and a rotation around their optical axis. We stated the important assumptions allowing to design the calibration method independently on the resolution of a sensor chip and to recover metrically calibrated camera by auto-calibration. Both, the epipolar geometry and the projection equation can be formulated on calibrated vectors even though the ambiguity in the auto-calibration. We discussed measures of error in epipolar geometry estimation to be able to mark outliers/inliers in estimation technique based on RANSAC strategy. Finally, we showed that the 3D reconstruction built from the calibrated vectors is a metric reconstruction.

5 Auto-calibration from Epipolar Geometry

Auto-calibration methods calibrate cameras from image correspondences without knowing the structure of observed scene a priori. Besides rigidity, no assumptions about the scene are made (e.g., presence of a calibration object or a special structure of the scene). After the auto-calibration, we will be able to compute 3D rays corresponding to image points. It means that angles between the rays will be correctly recovered and thus it will be possible to build 3D metric reconstructions.

First, we will show a radial distortion model [12] for narrow-angle view cameras which allows to formulate the camera auto-calibration as the Polynomial Eigenvalue Problem (PEP) [1, 91]. We took over and modified the Fitzgibbon's idea [30] of the auto-calibration method based on solving the PEP to omnidirectional cameras. Our approach is "stronger". It allows to obtain simultaneously a calibrated camera and an essential matrix in contrary to Fitzgibbon's method [30] where only the partially calibrated camera and the fundamental matrix are estimated. We derived auto-calibration methods, described in separate sections, for some types of catadioptric and dioptric omnidirectional cameras.

The proposed method for the auto-calibration of catadioptric cameras relies on the assumption that a full circular (or elliptical due to non-square pixels of the camera) field of view of the mirror is seen in the image. An ellipse can be fitted to the boundary of the view field in the acquired image and used to determine the image of the optical axis and thus the symmetry of projection in the image. This assumption holds only if the mirror axis is aligned with the camera optical axis. Fitzgibbon [30] assumed that a perspective camera was calibrated up to its radial distortion and the focal length (i.e., aspect ratio and skew parameters were known or were negligible). We can assume completely uncalibrated camera, since the full field of view visible in the image can be used to pre-calibrate the camera.

The center of symmetry $(u_0, v_0)^\top$ will be estimated from the center of the elliptical (usually circular) field of view. Parameters α_u, α_v, and skew s in the calibration matrix [43] are computed from the transformation of the view field ellipse into a circle. Thus, the intrinsic parameters, i.e., the calibration matrix [43] of a camera, become known. Only the parameters $\mathbf{a}''$ of the mirror or lens, Equation 4.3, remain unknown and will be estimated by the auto-calibration method.

Since the epipolar geometry, see Section 4.4, holds for central omnidirectional cameras, it can be and will be employed in the proposed auto-calibration method.

5.1. Narrow angle view cameras

Radial distortion of a *narrow-angle* lens can be represented (knowing the principal point and assuming square pixels and zero skew) by the division model introduced in [12] and used by Fitzgibbon [30], as follows

$$\mathbf{u}_p = \frac{1}{1 + b\,\|\mathbf{u}_x\|^2}\mathbf{u}_x \quad \Rightarrow \quad r_p = \frac{r_x}{1 + b\,r_x^2}\,, \tag{5.1}$$

where $\mathbf{u}_x = (u,\, v)^\top$ is a point in the original image, $r_x = \sqrt{u^2 + v^2}$ is its radius with respect to the center of the distortion, $\mathbf{u}_p = (p,\, q)^\top$ ($_p$ denotes perfect) is a point in the undistorted image, $r_p = \sqrt{p^2 + q^2}$ is its radius and b is the parameter describing the radial distortion. Equation (5.1) represents a mapping from a distorted image to an undistorted one. The origin is located in the center of the image and it is assumed that the distortion is radially symmetric with respect to the origin.

Image points and undistorted image points, expressed in homogeneous coordinates, represent one-dimensional subspaces in $\mathbb{R}^3$ (points in $\mathbb{P}^2$) emanating from a camera center $\mathbf{C}$, see vectors $\mathbf{v}_x$ and $\mathbf{v}_p$ in Figure 5.1. It holds for the angle θ between the vectors and the optical axis that

$$\tan\theta = \frac{r_p}{f} = \frac{r_x}{w}\,. \tag{5.2}$$

Substituting r_p from Equation (5.1) into Equation (5.2) leads to

$$w = f(1 + b\,r_x^2)\,.$$

The undistorted vector $\mathbf{v}_p$ computed from the vector $\mathbf{v}_x$ can be expressed as

$$\mathbf{v}_p = \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} u \\ v \\ f(1 + b\,r_x^2) \end{pmatrix} = \begin{pmatrix} u \\ v \\ f \end{pmatrix} + f\,b \begin{pmatrix} 0 \\ 0 \\ r_x^2 \end{pmatrix}\,. \tag{5.3}$$

Notice that the radial distortion parameter and the camera focal length f has to be known to compute the vectors $\mathbf{v}_p$ from image points $\mathbf{u}_x$.

The undistorted ray $\mathbf{v_p}$ can be projected back into the image plane by multiplying $\mathbf{v_p}$ by scalar $\frac{1}{1+br_x^2}$. The following holds

$$\mathbf{v}_p = \begin{pmatrix} u \\ v \\ f(1 + b\,r_x^2) \end{pmatrix} \simeq \begin{pmatrix} \frac{u}{1+b\,r_x^2} \\ \frac{v}{1+b\,r_x^2} \\ f \end{pmatrix}\,.$$

The first two coordinates of the multiplied vector $\mathbf{v_p}$ are now the coordinates of the undistorted point in the image plane with the radius

$$r_p = \sqrt{\left(\frac{u}{1 + b\,r_x^2}\right)^2 + \left(\frac{v}{1 + b\,r_x^2}\right)^2}\,, \tag{5.4}$$

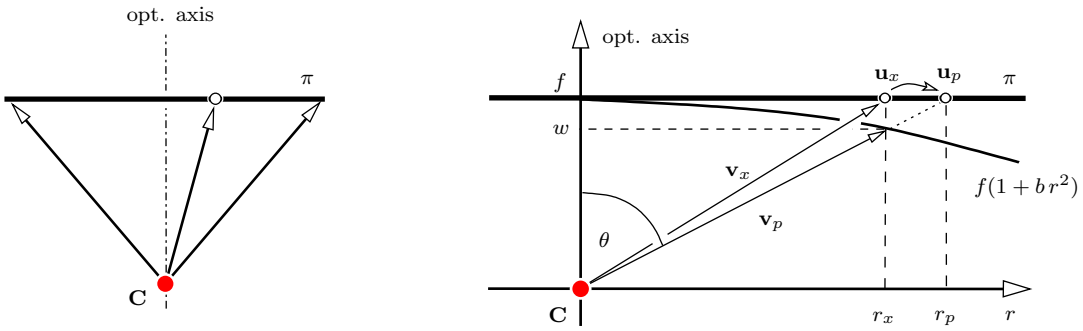
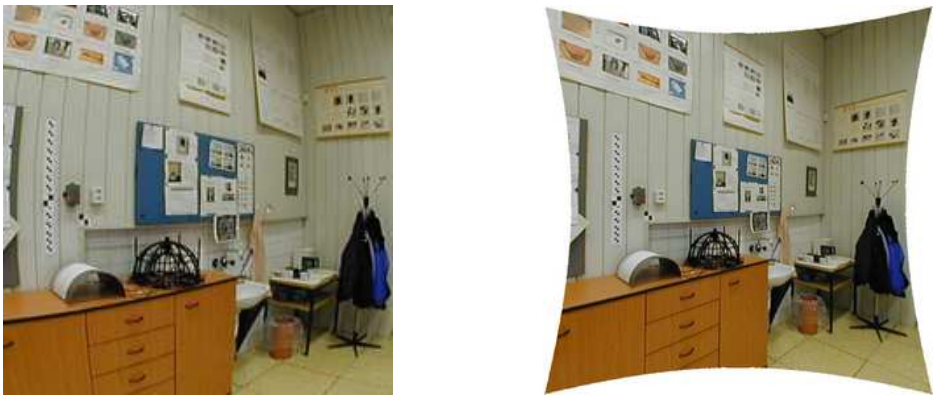


Figure 5.1.: A cross section of the projection plane π by a perpendicular plane passing the optical axis. (a) The perspective model of a narrow-angle view camera. (b) The geometrical interpretation of the elimination of radial distortion and the principle of creating 3D vector from image coordinates for a standard narrow-angle lens. r_x is the radius of a distorted image point, r_p is the radius of the corresponding undistorted image point, π is the image plane, $\mathbf{v_x} = (u, v, f)^\top$ and $\mathbf{v_p} = (u, v, w)^\top$ are the directional vectors of rays going through image points, f is the focal length of the lens, θ is the angle between the undistorted vector and the optical axis, $\mathbf{C}$ is the optical center.



(a) (b)

Figure 5.2.: (a) A narrow view-angle image with significant radial distortion. (b) The undistorted image obtained by the Fitzgibbon's method [30].

which is the same as in Equation (5.1). Notice that the change of the focal length does not affect the radius r_p and thus the undistorted image point. It follows that the focal length does not have to be known to eliminate the radial distortion in the image. However, the focal length has to be known to compute the corresponding 3D rays and thus to obtain the calibrated camera.

Provided that we want to eliminate the radial distortion in the image and not to calibrate the camera we can choose $f = 1$ and just estimate the parameter b. Then, Equation (5.3) is the same as in [30], and the relationship between original image point $\mathbf{u}_x = (u, v)^\top$ and undistorted image point $\mathbf{u}_p = (p, q)^\top$ is

$$\begin{pmatrix} p \\ q \\ 1 \end{pmatrix} = \alpha \left[\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} + b \begin{pmatrix} 0 \\ 0 \\ \|\mathbf{u}_x\|^2 \end{pmatrix} \right], \qquad \alpha \in \mathbb{R} \setminus \{0\}, \tag{5.5}$$

what can directly be substituted into the epipolar constraint, Equation (4.18), and rearranged to obtain the following equation

$$(\mathtt{D}_1 + b\,\mathtt{D}_2 + b^2\,\mathtt{D}_3)\,\mathbf{f} = \mathbf{0}, \tag{5.6}$$

known as the Polynomial Eigenvalue Problem (PEP) [1, 91]. Matrices $\mathtt{D}_i$ are the matrices composed of coordinates of point correspondences. The vector $\mathbf{f}$ is composed of elements of the fundamental matrix $\mathtt{F}$. The reader is refered to Fitzgibbon's paper [30] where more details can be found.

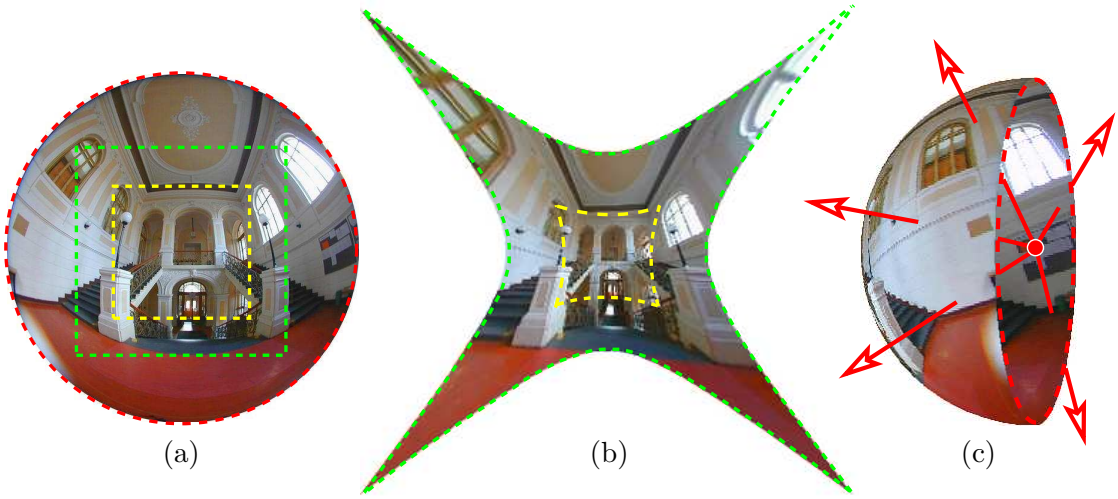
The division model, Equation (5.1), often gives better results than the standard polynomial models [30, 90] and allows to estimate parameter b and the fundamental matrix simultaneously from point correspondences between a pair of images by solving the PEP [30]. By using the estimated parameter b and Equation (5.1), the original image can be undistorted to a perspective one where all lines are indeed straight, see Figure 5.2. Since the parameter b is estimated and the image is undistorted using this b, the calibration matrix after the calibration becomes

$$\mathtt{K} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where f is the focal length in pixels. Since the focal length f is not known, the camera becomes only *partially* calibrated. As it was shown before, we do not need to know f to undistort the image and to determine the fundamental matrix $\mathtt{F}$ [30].

What would happen if the model, Equation (5.5), was used for omnidirectional cameras? Assume, e.g., that images were acquired by the Nikon FC–E8 fish-eye lens with field of view 183° and let us cut out only part of the image that corresponds to the field of view of a standard camera, see Figure 5.3. When increasing the view angle, two problems appear:

i) The projection of undistorted vectors back into the image plane for large view angles is restricted since the size of the undistorted image goes to infinity for the



(a) (b) (c)

Figure 5.3.: (a) Original image acquired by the Nikon FC–E8 fish-eye converter mounted on the Nikon COOLPIX digital camera with resolution 1600×1200 pixels. Yellow (smaller) and green (larger) dashed rectangles depict standard perspective cameras with the field of view 90° and 120°, respectively. The red circle depicts the field of view 183°. (b) The size of the undistorted images using Fitzgibbon's approach [30] goes to infinity for the field of view approaching 180°. (c) The complete field of view can be represented on a sphere.

field of view approaching 180°. The last coordinate of $\mathbf{v}_p$ tends to zero and finally there will be no intersection of the $\mathbf{v}_p$ with the image plane. Notice in Figure 5.3(b) that for 120° angle of view the corners of the undistorted image are significantly stretched.

ii) The function $r_p = k(r_x)$ in Equation (5.5) does not sufficiently capture the non-linear distortion for large angle of views. Moreover, if we represent the undistorted image by the spherical model, see Figure 5.3(c), it is necessary to investigate the non-linearity $\theta = k(r_x)$ instead of $r_p = k(r_x)$.

In the next, we will show how the Fitzgibbon's method [30] can be generalized to some omnidirectional cameras and how the above two problems can be solved. For each camera,

- the camera model is shown,
- the functions $m(.)$, $i(.)$ from Theorem 1 are formulated and it is proved that the calibration method can be designed independently on the sensor chip resolution,
- the calibrated vectors $\mathbf{p}$ are adjusted (multiplied or linearized) in such a way that the epipolar constraint leads to the PEP estimation problem,
- an experiment showing the performance of the method is presented.

5.2. Para-catadioptric camera

The para-catadioptric camera (PCD), see Figure 5.5, is composed of a convex parabolic mirror and an orthographic camera [10]. The orthographic camera is assembled with the parabolic mirror so that the rays of the orthographic camera are parallel to the mirror symmetry axis, see Figure 5.4. An image point $\mathbf{u}$ is orthographically projected on the mirror and reflected such that the ray $\mathbf{p}$ passes through the focal point $\mathbf{F}$. All rays intersect in the $\mathbf{F}$ and therefore the PCD camera possesses the central projection.

5.2.1. Camera model

All coordinates will be expressed in a mirror Cartesian coordinate system placed in $\mathbf{F}$, see Figure 5.4, with z axis aligned with the axis of the mirror. Let us consider a paraboloid of revolution with the equation

$$z = \frac{a''^2 - \|\mathbf{u}''\|^2}{2a''}\,, \tag{5.7}$$

where $\mathbf{u}'' = (-u'',\, v'')^\top$ is a point in the sensor plane ($-u''$ since the image is mirrored), a'' is the parameter of the mirror. It is twice the distance from the vertex to the focal point and captures the shape of the parabolic mirror. It will be estimated during the calibration process.

Let us emphasize that symbols with the double prime, e.g., a'', $\mathbf{u}''$, stand for the entities measured in units of the sensor Cartesian coordinate system, usually in the metric system. The same symbols marked without the prime, e.g., a, $\mathbf{u}$, represent the same entities but measured in the Cartesian coordinate system of the pre-calibrated image in *pixels*.

The model of the PCD camera according to Equation (4.3) reads as

$$\mathbf{p}'' = \begin{pmatrix} \mathbf{u}'' \\ \frac{a''^2 - \|\mathbf{u}''\|^2}{2a''} \end{pmatrix}. \tag{5.8}$$

Thus, $h(\|\mathbf{u}''\|) = 1$ and $g(\|\mathbf{u}''\|) = \frac{a''^2 - \|\mathbf{u}''\|}{2a''}$ w.r.t. the model in Equation (4.3). The relationship between $\mathbf{u}''$ and $\mathbf{u}'$ is given by Equation (4.5). As it was explained in Section 4.3, the point $\mathbf{u}'$, measured in the digital image, is transformed to the point $\mathbf{u}$ in the pre-calibrated image by applying the affine transformation in Equation (4.7), i.e., by transforming a view field ellipse to a circle. The projection becomes radially symmetric and the projection Equation (4.6) holds.

The next result follows from Theorem 1 in Section 4.3.

Result 1 *Let the functions $m(\rho, \mathbf{a}'') = \frac{a''}{\rho}$ and $i(\rho) = 1$. Then, the Theorem 1 holds for the PCD cameras and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} = \begin{pmatrix} \mathbf{u} \\ \frac{a^2 - \|\mathbf{u}\|^2}{2a} \end{pmatrix}. \tag{5.9}$$

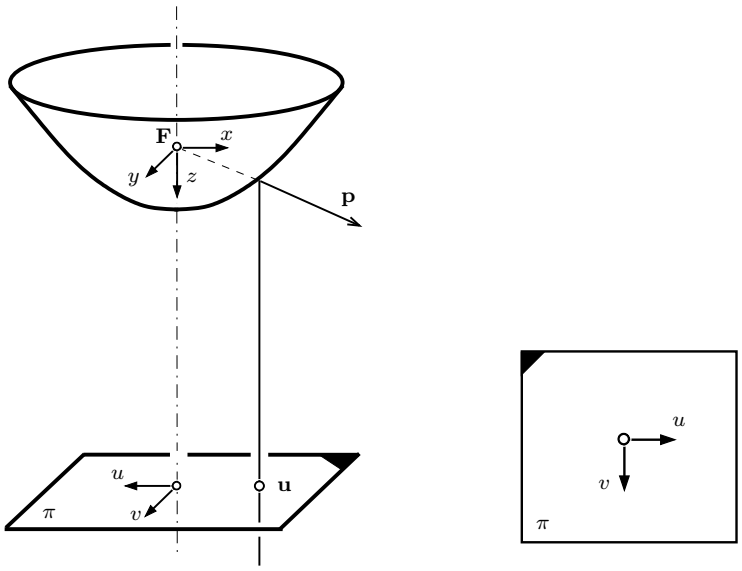


Figure 5.4.: Left: Coordinate system of the para-catadioptric camera. The origin is located in **F**. Right: A coordinate system in the digital image.

Figure 5.5.: A para-catadioptric camera consisting of an orthographic camera and a parabolic mirror.

Proof. Substituting $\mathbf{u}'' = \rho\,\mathtt{R}\mathbf{u}$ from Equation (4.9) into the Equation (5.8) leads to

$$\begin{aligned}\mathbf{p}'' &= \begin{pmatrix} \mathbf{u}'' \\ \frac{a''^2 - \|\mathbf{u}''\|^2}{2a''} \end{pmatrix} = \begin{pmatrix} \rho\,\mathtt{R}\mathbf{u} \\ \frac{a''^2 - \|\rho\,\mathtt{R}\mathbf{u}\|^2}{2a''} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\left(\frac{a''}{\rho}\right)^2 - \|\mathbf{u}\|^2}{2\frac{a''}{\rho}} \end{pmatrix} = \\ &= \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{a^2 - \|\mathbf{u}\|^2}{2a} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}\,. \end{aligned}$$

□

It follows from the proof of Result 1 that ρ is absorbed to the parameter a. The ρ converts a'' from millimetres to a in pixels. It means that if the camera is calibrated from $\mathbf{u}$ measured in the pre-calibrated image (in pixels), then the estimated parameter a will be in pixels as well. The ambiguity captured by ρ and $\mathtt{R}$ does not affect the angles between vectors $\mathbf{p}$ and the optical axis.

Therefore, we get the following consequence of Result 1.

Consequence 3 *The unknown scale ρ capturing the resolution of the camera sensor chip does not add an extra variable in the camera model and does not increase the number of unknown parameters which are estimated during the calibration process. It follows that the Consequence 1 and Consequence 2 in Section 4.3 hold and the calibration method can be designed independently on the camera chip resolution.*

5.2.2. Camera auto-calibration

By the calibration of the PCD camera we understand the determination of the matrix $\mathtt{A}$ and the vector $\mathbf{t}$, Equation (4.9) and the parameter a of the non-linear function g, so that all vectors, Equation (5.9), fulfill epipolar geometry, Equation (4.20), and the projection equation (4.6).

The vector $\mathbf{p}_1$ in the left image and $\mathbf{p}_2$ in the right image, Equation (5.9), can be substituted into the epipolar constraint, Equation (4.20), as follows

$$\begin{gathered} \mathbf{p}_2^\top \mathtt{F}\,\mathbf{p}_1 = 0\,, \\ \begin{pmatrix} -u_2 & v_2 & \frac{a^2 - \|\mathbf{u}_2\|^2}{2a} \end{pmatrix} \mathtt{F} \begin{pmatrix} -u_1 \\ v_1 \\ \frac{a^2 - \|\mathbf{u}_1\|^2}{2a} \end{pmatrix} = 0\,. \end{gathered} \tag{5.10}$$

Arranging and gathering the point coordinates and radii into five design matrices, we obtain the equation

$$\left(\mathtt{D}_1 + a\,\mathtt{D}_2 + a^2\mathtt{D}_3 + a^3\mathtt{D}_4 + a^4\mathtt{D}_5\right)\mathbf{f} = \mathbf{0}\,, \tag{5.11}$$

quartic (degree 4) in parameter a and linear in $\mathbf{f}$, known as the Polynomial Eigenvalue Problem (PEP) [1]. Efficient algorithms for solving the PEP are available [1], e.g., MATLAB solves the PEP by the function `polyeig`.

Every point correspondence adds one row into the matrices $\mathtt{D}_i$. The unknown vector $\mathbf{f}$ (created from the matrix $\mathtt{F}$) and one row from each of the five design matrices $\mathtt{D}_i$, composed of image point coordinates, have the following form:

$$\begin{aligned}
\mathbf{f} &= [\, F_{11},\, F_{12},\, F_{13},\, F_{23} \quad \ldots \quad F_{33} \,]^\top , \\
\mathtt{D}_1 &= [\, 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; r_2^2\, r_1 \,] , \\
\mathtt{D}_2 &= [\, 0 \;\; 0 \;\; 2u_2\, r_1^2 \;\; 0 \;\; 0 \;\; -2v_2\, r_1^2 \;\; 2u_1\, r_2^2 \;\; -2v_1\, r_2^2 \;\; 0 \,] , \\
\mathtt{D}_3 &= [\, 4u_2\, u_1 \;\; -4u_2\, v_1 \;\; 0 \;\; -4v_2\, u_1 \;\; 4v_2\, v_1 \;\; 0 \;\; 0 \;\; 0 \;\; -r_2^2 - r_1 \,] , \\
\mathtt{D}_4 &= [\, 0 \;\; 0 \;\; -2u_2 \;\; 0 \;\; 0 \;\; 2v_2 \;\; -2u_1 \;\; 2v_1 \;\; 0 \,] , \\
\mathtt{D}_5 &= [\, 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 1 \,] ,
\end{aligned} \tag{5.12}$$

where $r_1 = \|\mathbf{u}_1\|$ and $r_2 = \|\mathbf{u}_2\|$.

The matrices $\mathtt{D}_i$ should be square and therefore 9 point matches are needed for solving the PEP in Equation (5.11). There are 36 solutions for a. Many of them are zero, infinite or complex. In practice, no more than 4 solutions are considerable, see Appendix A.2. Out of these, the one which has the smallest angular error, Equation (4.22), is chosen.

The estimated parameter a allows to compute $\mathbf{p}_1$, $\mathbf{p}_2$, Equation (5.9), related to $\mathbf{p}_1''$, $\mathbf{p}_2''$ up to a scale and a rotation around the optical axis. However, the angles between the vectors and the optical axis are correctly recovered. Therefore, as a result of solving the PEP, Equation (5.11), we get an essential matrix (instead of a fundamental one) [43]. It allows us to determine the relative rotation and the translation direction of a pair of PCD cameras.

Automatically established point correspondences are usually contaminated by mismatches. Since the auto-calibration method needs 9 points, it can be easily incorporated in a 9-point RANSAC estimation technique to handle outliers in point matches. The computational complexity increases twice for 50% contamination by outliers in comparison to the standard 8-point RANSAC algorithm [43] for the fundamental matrix computation. It will be described later in Section 5.6 how to speed up the convergence of the RANSAC by using a bucketing technique.

As the result of the auto-calibration method, the calibrated camera, the estimated essential matrix, and the outliers in point matches are obtained, see Figure 5.6. Results can be directly used to obtain a 3D metric reconstruction from two omnidirectional images.

5.2.3. Experiment

We mounted a PCD camera, see Figure 5.5, on a turntable such that the trajectory of its optical center (the mirror focal point $\mathbf{F}$) was circular, see Figure 5.7. The PCD camera was set up from the digital camera Canon EOS–1Ds with the diameter of view field circle equal 2500 pixels, and the commercially available parabolic mirror [52] with

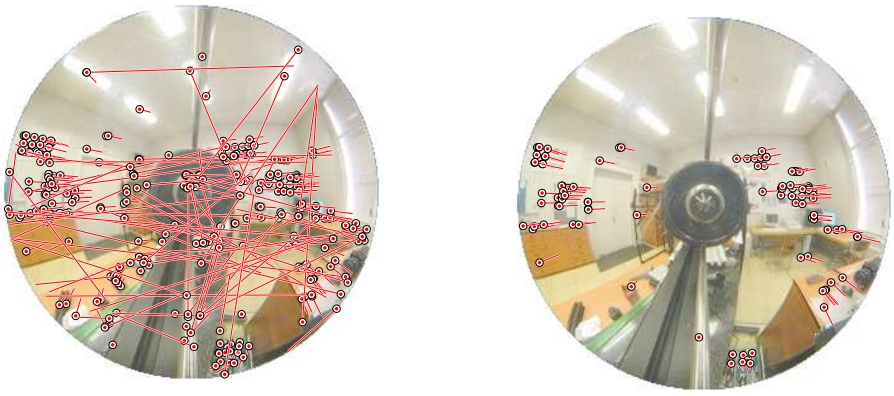


Figure 5.6.: Outliers detection during the calibration process for a PCD camera. Left: An acquired digital image from a stereopair with all tentative correspondences found by [58]. Circles mark points in the first image, lines join their matches in the next one. Right: The same image with the validated inliers only.

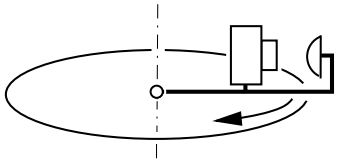
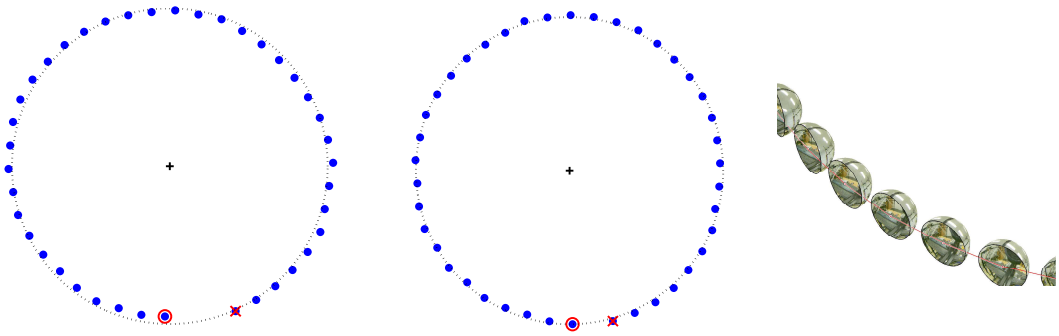
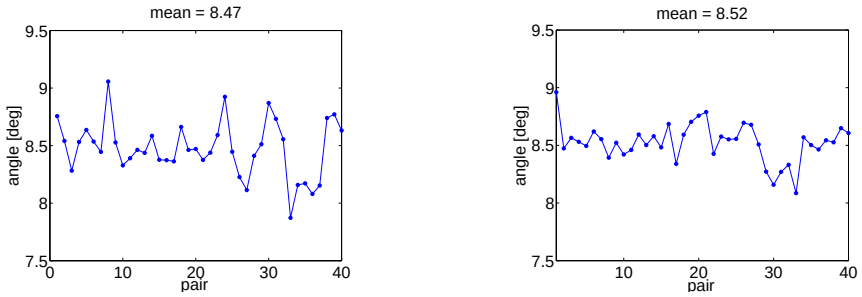


Figure 5.7.: An experiment setup of a circular sequence with a PCD camera.



(a) (b) (c)

Figure 5.8.: Estimated trajectories from the circle sequence. Red ∘ depicts the starting position, × depicts the end position. (a) The trajectory estimated during the calibration. (b) The trajectory estimated after calibration using the same $\hat{a}$ for each pair. (c) A detail view of the estimated camera orientations in the circle sequence.



(a) (b)

Figure 5.9.: Estimated angles of rotation between two consequent camera positions in the plane of rotation (containing all PCD camera centers) for every consecutive pair (a) using different a's for each image pair, and (b) using the same $\hat{a}$ for whole sequence.

equation in a metric system

$$z_{mm} = \frac{\sqrt{x_{mm}^2 + y_{mm}^2}}{39.22}.$$

Images were acquired every 9°, 40 images in total. The correspondences for every consecutive pair of images were established by the wide baseline stereo technique [58] and used in the calibration process based on the RANSAC with bucketing (described in Section 5.6).

An essential matrix `F`, the camera parameter a, and correct point matches, see Figure 5.6, were obtained between every consecutive pair as a result of the calibration method. From the essential matrix `F`, a relative camera rotation and a direction of the camera translation can be computed [43]. For obtaining the magnitudes of the translational vectors we would need to reconstruct the observed scene. It was not the task of this experiment. Instead, we normalized the translational vectors to have unit length. The final trajectory of the optical center can be seen in Figure 5.8(a). The estimated angles of rotations are shown in Figure 5.9(a).

After calibrating all consecutive image pairs, the parameter $\hat{a}$ was computed as the mean over all computed a's. We used the estimated $\hat{a}$ to create 3D rays corresponding to the point correspondences, Equation (5.9), and applied the standard linear 8-point algorithm [43] to re-compute the essential matrices. See Figure 5.8 and Figure 5.9 to compare the results.

It can be seen from the results that the estimation of the camera parameter $\hat{a}$ from more pairs of images leads to better accuracy of the estimated trajectory. However, the end position does not approach the starting one, neither using the various a's nor using the one $\hat{a}$ for entire sequence, as it should. The error accumulated over the sequence reaches 18° at the end, what is 5% relative error. There is an offset, approximately 0.5°, on each pair in the angle estimation. The directions of the cameras are correctly estimated during whole sequence, see Figure 5.8(c).

The error in the trajectory estimation arised since a non-ideal orthographic camera

was used. It is difficult to manufacture an ideal orthographic camera, i.e., to have all rays parallel. The real orthographic cameras usually work only approximately in some range of distances. It causes that the PCD camera becomes slightly non-central and a more complicated model has to be used to obtain more precise SFM. The real non-central catadioptric cameras will be described later in Chapter 6 and it will be shown how the accuracy of SFM increases.

Nevertheless, the tentative correspondences were correctly validated by the central model, see Figure 5.6. It shows the strength of the method. It is possible, and advisable, to validate the tentative matches using the central model as a starting point for a non-linear bundle adjustment with more accurate model.

5.3. Hyperbolic catadioptric camera

An ideal hyperbolic catadioptric (HCD) camera, see Figure 5.11, is properly assembled from a hyperbolic mirror and a perspective camera with the camera optical center placed in the second focal point of the hyperboloid. Such a HCD camera realizes a central projection, hence it possesses a single viewpoint, see Figure 4.4(b) and Figure 5.10. An image point $\mathbf{u}$ is perspectively projected on the mirror and reflected such that the ray $\mathbf{p}$ passes through the focal point $\mathbf{F}$. All rays intersect in $\mathbf{F}$ and therefore the HCD camera possesses the central projection.

5.3.1. Camera model

All coordinates will be expressed in the mirror Cartesian coordinate system placed in $\mathbf{F}$, see Figure 5.10, with z axis aligned with the axis of the mirror. Let us consider the hyperboloid of revolution with the equation

$$\frac{(z-e'')^2}{a''^2} - \frac{x^2+y^2}{b''^2} = 1\,, \tag{5.13}$$

where a'', b'' are parameters of the hyperbolic mirror and $e'' = \sqrt{a''^2 + b''^2}$ is its eccentricity. The parameters capture the shape of the hyperbolic mirror and will be estimated during the calibration process. The model of the HCD camera according to Equation (4.3) becomes

$$\mathbf{p}'' = a'' \left[\frac{c''^2 \left(\sqrt{1+c''^2} + \sqrt{\frac{\|\mathbf{u}''\|^2}{f''^2}+1} \right)}{c''^2 - \frac{\|\mathbf{u}''\|^2}{f''^2}} \begin{pmatrix} -\frac{u''}{f''} \\ \frac{v''}{f''} \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 2\sqrt{1+c''^2} \end{pmatrix} \right], \tag{5.14}$$

where $c'' = \frac{b''}{a''}$ is a new mirror parameter composed of a'' and b'', f'' is the focal length of a perspective camera, $\mathbf{u}'' = (-u'',\, v'')^\top$ is a point measured in the sensor Cartesian coordinate system. The central model is a special case of the non-central one. The derivation of Equation (5.14) from Equation (5.13) will become more clear in Section 6.2, where

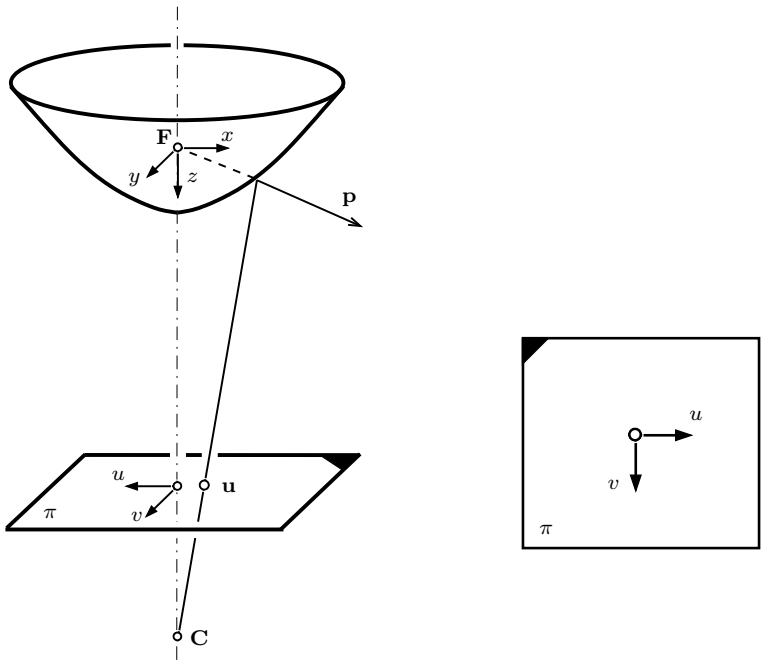


Figure 5.10.: Left: Coordinate system of the hyperbolic catadioptric camera. The origin is located in **F**. Right: A coordinate system in the digital image.

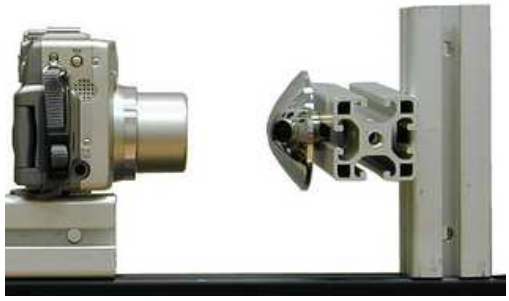


Figure 5.11.: A hyperbolic catadioptric camera consisting of a perspective camera and a hyperbolic mirror.

the full non-central model of HCD cameras will be derived. We need the direction of the vector $\mathbf{p}''$, not the magnitude, therefore, a'' before the square brackets in Equation (5.14) can be omitted. Then, the direction of the vector $\mathbf{p}''$ is a *two*-parametric function of parameters c'' and f''. The functions g, h, in notation of Equation (4.3), are as follows

$$\begin{aligned} h(\|\mathbf{u}''\|) &= \frac{1}{f''} \frac{c''^2 \left(\sqrt{1+c''^2} + \sqrt{\frac{\|\mathbf{u}''\|^2}{f''^2}+1} \right)}{c''^2 - \frac{\|\mathbf{u}''\|^2}{f''^2}}, \\ g(\|\mathbf{u}''\|) &= 2\sqrt{1+c''^2} - f'' \, h(\|\mathbf{u}''\|). \end{aligned} \tag{5.15}$$

The focal length f'' of the perspective camera is a priori unknown. Suppose that the ratio $K = \frac{height}{radius} = \frac{H''}{R''}$ of the mirror is known, see Figure 5.12. The ratio can be obtained easier (by measuring the mirror) than the camera focal length f''. From Equation (5.13) and $K = \frac{H''}{R''}$ it holds that the radius

$$R'' = -\frac{2K \, a'' b''^2}{K^2 \, b''^2 - a''^2}. \tag{5.16}$$

Assuming the perspective projection, see Figure 5.12, it follows that

$$\frac{2e'' + K \, R'' - (e'' - a'')}{R''} = \frac{f''}{r''_{max}}, \tag{5.17}$$

where r''_{max} is the radius of the view field circle in the sensor plane, see Figure 5.13. Substituting Equation (5.16) into Equation (5.17) yields

$$f'' = -\frac{r''_{max}(K^2 c''^2 - 1)}{2K c''^2} \left(\sqrt{1+c''^2} - \frac{2K^2 c''^2}{K^2 c''^2 - 1} + 1 \right). \tag{5.18}$$

Equation (5.18) can be substituted into the Equation (5.14) to obtain the 3D vector $\mathbf{p}''$ as a *one*-parametric function of the parameter c''. The c'' will be estimated by a calibration process. The ratio K is supposed to be known.

Result 2 *Let the functions $m(\rho, \mathbf{a}'') = c''$, and $i(\rho) = \frac{a''}{\rho}$. Then, the Theorem 1 holds for the HCD cameras and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} \simeq \frac{c^2 \left(\sqrt{1+c^2} + \sqrt{\frac{\|\mathbf{u}''\|^2}{f^2}+1} \right)}{c^2 - \frac{\|\mathbf{u}\|^2}{f^2}} \begin{pmatrix} -\frac{u}{f} \\ \frac{v}{f} \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 2\sqrt{1+c^2} \end{pmatrix}, \tag{5.19}$$

where

$$f = -\frac{r_{max}(K^2 c^2 - 1)}{2K c^2} \left(\sqrt{1+c^2} - \frac{2K^2 c^2}{K^2 c^2 - 1} + 1 \right). \tag{5.20}$$

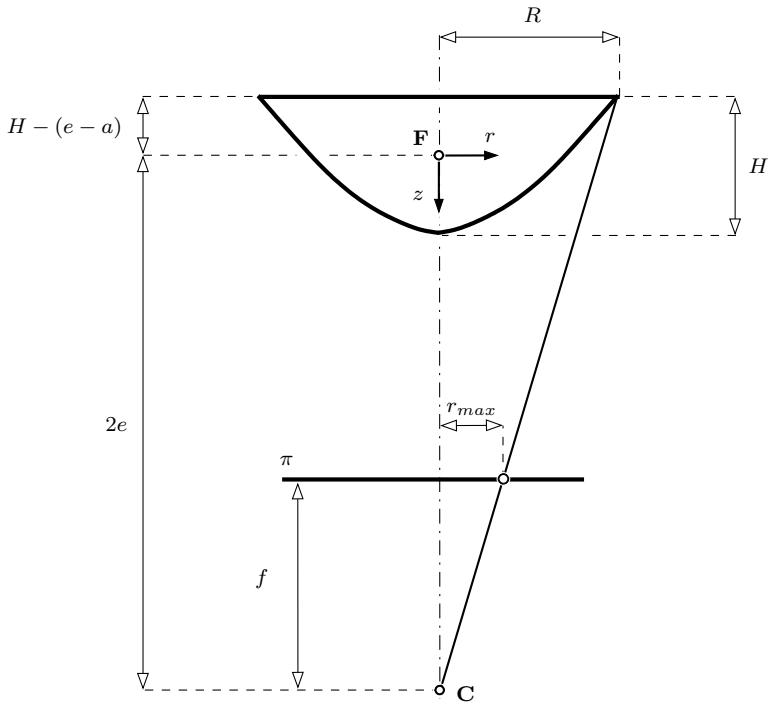


Figure 5.12.: Schematic diagram of a hyperbolic catadioptric camera cross section with labelled symbols used in the Section 5.3.1.

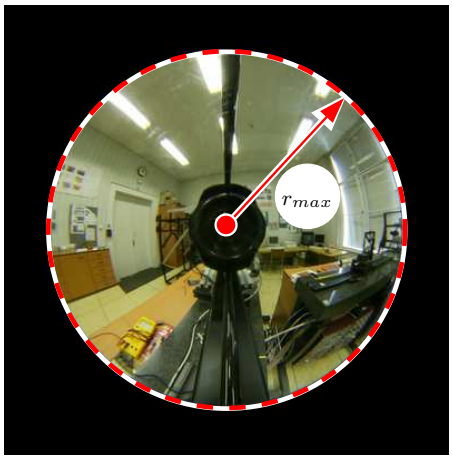


Figure 5.13.: A pre-calibrated image with a circular field of view. The radius r_{max} is obtained as the radius of the view field circle.

Proof. Substituting $\mathbf{u}'' = \rho\,\mathtt{R}\mathbf{u}$ into the Equation (5.14) leads to

$$\mathbf{p}'' = a''\begin{pmatrix}\mathtt{R} & \\ & 1\end{pmatrix}\left[\frac{c''^2\left(\sqrt{1+c''^2}+\sqrt{\frac{\|\mathbf{u}\|^2}{\frac{f''^2}{\rho^2}}+1}\right)}{c''^2-\frac{\|\mathbf{u}\|^2}{\frac{f''^2}{\rho^2}}}\begin{pmatrix}\frac{\mathbf{u}}{\frac{f''}{\rho}}\\ -1\end{pmatrix}+\begin{pmatrix}\mathbf{0}\\ 2\sqrt{1+c''^2}\end{pmatrix}\right] =$$

$$= a''\begin{pmatrix}\mathtt{R} & \\ & 1\end{pmatrix}\left[\frac{c^2\left(\sqrt{1+c^2}+\sqrt{\frac{\|\mathbf{u}\|^2}{f^2}+1}\right)}{c^2-\frac{\|\mathbf{u}\|^2}{f^2}}\begin{pmatrix}\frac{\mathbf{u}}{f}\\ -1\end{pmatrix}+\begin{pmatrix}\mathbf{0}\\ 2\sqrt{1+c^2}\end{pmatrix}\right] = a''\begin{pmatrix}\mathtt{R} & \\ & 1\end{pmatrix}\mathbf{p}\,,$$

where

$$f = \frac{f''}{\rho} = -\frac{\frac{r''_{max}}{\rho}(K^2c''^2-1)}{2Kc''^2}\left(\sqrt{1+c''^2}-\frac{2K^2c''^2}{K^2c''^2-1}+1\right) =$$

$$= -\frac{r_{max}(K^2c^2-1)}{2Kc^2}\left(\sqrt{1+c^2}-\frac{2K^2c^2}{K^2c^2-1}+1\right).$$

□

It follows from the above proof that the unknown scale ρ is absorbed into the radius r_{max} of the view field circle. The ρ converts the r''_{max} (in millimetres) to r_{max} (in pixels). The radius r_{max} can be obtained as the radius of the view field circle in the pre-calibrated image since we measure there in pixels, see Figure 5.13. The parameter c'' is unitless variable. The change of units does not affect it, hence $c'' = c$. The proof has shown that Consequence 3 in Section 5.2.1 also holds for HCD cameras.

5.3.2. Camera auto-calibration

The vector $\mathbf{p}_1$ in the left image and $\mathbf{p}_2$ in the right image, Equation (5.19), can be substituted into the epipolar constraint $\mathbf{p}_2^\top\mathtt{F}\,\mathbf{p}_1 = 0$ in the same manner as it was done for the PCD camera. Unlike in the case of the PCD camera, it does not lead directly to the PEP here. In order to obtain the PEP (and thus to set a direct solution of $\mathtt{F}$ and c), a linearization of the vector $\mathbf{p}$ is necessary. The vector $\mathbf{p}$, Equation (5.19), will be linearized in c_0 using the Taylor series

$$\mathbf{p} \approx \tilde{\mathbf{p}} = \begin{pmatrix}p_x\\ p_y\\ p_z\end{pmatrix}_{c=c_0} + (c-c_0)\begin{pmatrix}\dot{p}_x\\ \dot{p}_y\\ \dot{p}_z\end{pmatrix}_{c=c_0} =$$

$$= \begin{pmatrix}p_x-\dot{p}_xc_0\\ p_y-\dot{p}_yc_0\\ p_z-\dot{p}_zc_0\end{pmatrix}_{c=c_0} + c\begin{pmatrix}\dot{p}_x\\ \dot{p}_y\\ \dot{p}_z\end{pmatrix}_{c=c_0} =$$

$$= \begin{pmatrix}p\\ q\\ w\end{pmatrix} + c\begin{pmatrix}s\\ t\\ o\end{pmatrix} = \mathbf{x} + c\,\mathbf{s}\,, \tag{5.21}$$

where vectors $\mathbf{x}$ and $\mathbf{s}$ can be computed directly from known image coordinates of the corresponding points, and $\dot{p}_i$ stands for $\frac{\partial p_i}{\partial c}$. We used MAPLE [49] to obtain analytical forms of $\frac{\partial p_i}{\partial c}$. The forms are too extensive to be listed here, see appendix A.3.

The epipolar constraint, Equation (4.20), can be written as

$$\begin{aligned} \tilde{\mathbf{p}}_2^\top \mathtt{F}\, \tilde{\mathbf{p}}_1 &= 0\,, \\ (\mathbf{x}_2 + c\,\mathbf{s}_2)^\top \mathtt{F}\, (\mathbf{x}_1 + c\,\mathbf{s}_1) &= 0\,. \end{aligned} \tag{5.22}$$

Gathering the point coordinates into three design matrices, we obtain the equation

$$\left(\mathtt{D}_1 + c\,\mathtt{D}_2 + c^2\,\mathtt{D}_3\right)\mathbf{f} = \mathbf{0}\,, \tag{5.23}$$

quadratic in the parameter c and linear in $\mathbf{f}$, known as the Quadratic Eigenvalue Problem (QEP) [1, 91].

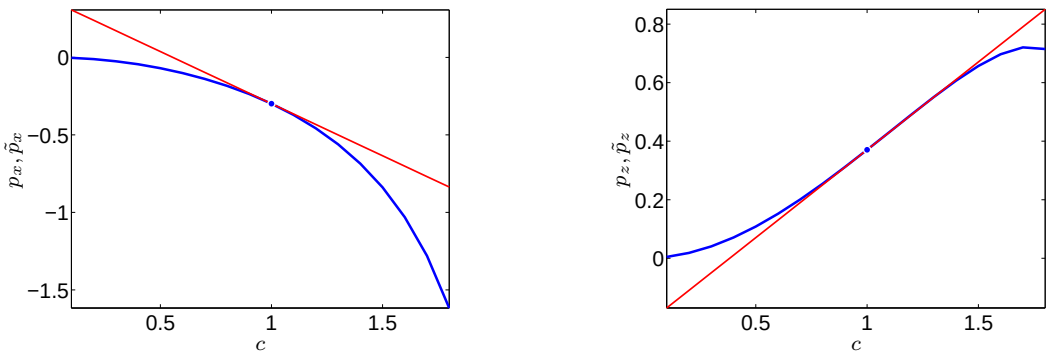
Every point correspondence adds one row into the matrices $\mathtt{D}_i$. The unknown vector $\mathbf{f}$ (created from the matrix $\mathtt{F}$) has the form as in Equation (5.12) and one row of each of the three design matrices $\mathtt{D}_i$, composed of image point coordinates, have the following form:

$$\begin{aligned} \mathtt{D}_1 &= \left[\; p_2p_1 \;\; p_2q_1 \;\; p_2w_1 \;\; q_2p_1 \;\; q_2q_1 \;\; q_2w_1 \;\; w_2p_1 \;\; w_2q_1 \;\; w_2w_1 \;\right], \\ \mathtt{D}_2 &= \left[\; p_2s_1 + s_2p_1 \;\; p_2t_1 + s_2q_1 \;\; p_2o_1 + s_2w_1 \;\; q_2s_1 + t_2p_1 \;\; q_2t_1 + t_2q_1 \;\; \right. \\ &\qquad \left. q_2o_1 + t_2w_1 \;\; w_2s_1 + o_2p_1 \;\; w_2t_1 + o_2q_1 \;\; w_2o_1 + o_2w_1 \;\right], \\ \mathtt{D}_3 &= \left[\; s_2s_1 \;\; s_2t_1 \;\; s_2o_1 \;\; t_2s_1 \;\; t_2t_1 \;\; t_2o_1 \;\; o_2s_1 \;\; o_2t_1 \;\; o_2o_1 \;\right]. \end{aligned} \tag{5.24}$$

The matrices $\mathtt{D}_i$ have to be square. Therefore, the minimum number of correspondences is 9 and the solution is obtained by $\mathtt{polyeig}(\mathtt{D}_3, \mathtt{D}_2, \mathtt{D}_1)$. If more than 9 points are to be used, the least-squares solution can be obtain by left-multiplication by $\mathtt{D}_i^\top$. It does not change the solution, but allows square solvers to be used. The solution is obtained by $\mathtt{polyeig}(\mathtt{D}_1^\top\mathtt{D}_3, \mathtt{D}_1^\top\mathtt{D}_2, \mathtt{D}_1^\top\mathtt{D}_1)$.

As a result of solving Equation (5.23), there are 18 solutions for c and the corresponding $\mathtt{F}$'s. Many of the solutions are zero, infinite or complex. In practice no more than 4 solutions are considerable, see Appendix A.2. Out of these, the one which has the smallest angular error, Equation (4.22), is chosen.

We end up with the same auto-calibration method as for PCD cameras. There are disadvantages of using HCD cameras compared to PCD cameras. First problem is that the ratio height/radius of the mirror has to be known to achieve the same computational complexity of the auto-calibration method as for the PCD cameras, i.e., to arrive at a 9-point RANSAC. Second, a point of linearization c_0 has to be somehow chosen. However, our experiments show that the method is not very sensitive to the choice of the c_0. The initial estimate $c_0 = 1$ was sufficient in all our experiments. The original vector $\mathbf{p}$ and its linearization is shown in Figure 5.14.



(a) (b)

Figure 5.14.: Plot of the first (a) and the third (b) coordinate of the vector $\mathbf{p}$ from Equation (5.19) and its linearization $\tilde{\mathbf{p}}$ in $c_0 = 1$ for image point $\mathbf{u} = (100, 0)^\top$. Blue thick curves are original functions, red thin lines are the linear approximations. Blue dots represent c_0, i.e., the points of the linearization.

5.3.3. Experiment

We performed the same circular experiment as for the PCD camera as described in Section 5.2.3. The HCD camera, Figure 5.11, was composed from the digital Canon PowerShot G2 with the diameter of the view field circle equal 1700 pixels and a commercially available hyperbolic mirror [51] with equation (in a metric system)

$$\frac{z_{mm}^2}{789.3} - \frac{x_{mm}^2 + y_{mm}^2}{548.1} = 1.$$

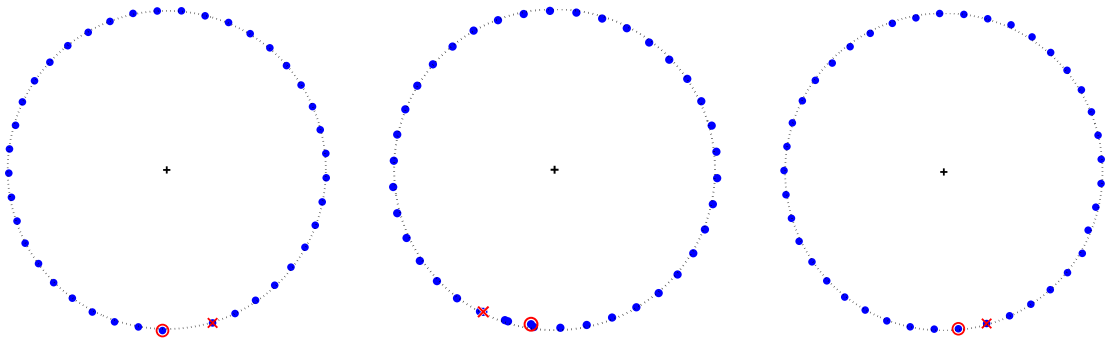
We roughly aligned the camera center to the second (outer) mirror focus in order to obtain a central camera. Images were acquired every 9°, 40 images in total. The correspondences for every consecutive pair of images were established by the wide baseline stereo technique [58] and used in the calibration process based on the RANSAC.

We used the same procedure as in the PCD camera experiment in Section 5.2.3. First, we used the different c's for each image pair, see Figure 5.16(a). Second, we used one $\hat{c}$ for the whole sequence, see Figure 5.16(c). Refer to Section 5.2.3 for more details.

It can be seen from the results, as for the PCD camera, that the estimation of the camera parameter from more pairs of images leads to better accuracy of the estimated trajectory. The end position does not approach the starting one, neither with the first (different c's) nor with the second approach (one $\hat{c}$), as it should. The error accumulated over the sequence reaches 10° at the end, what is 2.8% relative error. There is an offset, approximately 0.25°, on each pair in the angle estimation.

The error in trajectory estimation arised due to the same reason as for the PCD camera, i.e., the mirror is not aligned with the camera absolutely precisely and becomes slightly non-central. The real non-central HCD catadioptric camera model will be derived later in Chapter 6. The experiments show that the HCD camera is (twice) more

Figure 5.15.: Outliers detection during the calibration process for the HCD camera. Left: An acquired digital image from a stereopair with all tentative correspondences found by [58]. Circles mark points in the first image, lines join their matches in the next one. Right: The same image with the validated inliers only.



(a) (b) (c)

Figure 5.16.: Trajectory estimation for a circle sequence. Red ∘ depicts the starting position, × depicts the end position. (a) The trajectory estimated during the calibration. (b) The estimated trajectory using PCD model. (c) The trajectory estimated after the calibration using the same $\hat{c}$ for each pair.

accurate than the PCD camera. The reason is that the HCD camera was better aligned and thus closer to a central camera.

We tried to apply the auto-calibration method designed for PCD cameras to the previous circular sequence (acquired by the HCD camera) to see the behaviour of the auto-calibration method. It can be seen from Figure 5.16(b) that the estimated trajectory was approximately correct, however, 20° error appeared at the end. It follows that the PCD camera model roughly approximates the HCD camera. Therefore, it is possible to use the auto-calibration method for a rough estimate of the SFM and validation of the point matches if the mirror is close, though not exactly correct, to the shape for which the method is designed.

5.4. Spherical catadioptric camera

The spherical catadioptric camera (SCD), see Figure 5.18, consists of a spherical mirror and a perspective camera [10]. Compared to other quadric mirrors, e.g., hyperbolic, parabolic, or elliptical, the spherical mirror can be most easily manufactured. Other advantages of the spherical mirror are that it provides images with fewer blur and the equation of the surface is simple (one-parametric). The spherical mirror is a surface that does not possess a single projection center when observed by any central (perspective or orthographic) camera. However, a central model can be found that is sufficiently simple and accurate enough to distinguish between correct (inliers) and incorrect (outliers) point matches and which can be used to get an initial estimate of the camera motion.

The disadvantage of the spherical catadioptric camera is that the field of view can be only slightly changed by varying the distance of a perspective camera above the mirror.

5.4.1. Camera model

The natural non-central SCD camera model will be approximated by a central one. Suppose a fictive optical center $\mathbf{F}_f$ moved by $k\,a''$ from the center of the sphere, see Figure 5.19. All coordinates will be expressed in the mirror Cartesian coordinate system placed in $\mathbf{F}_f$, see Figure 5.17, with z axis passing through the camera and the sphere centers. The mirror sphere has the equation

$$x^2 + y^2 + (z + k\,a'')^2 = a''^2\,, \tag{5.25}$$

where a'' is the radius of the sphere. The model of the SCD camera according to Equation (4.3) is as follows

$$\mathbf{p}'' = a'' \left[\frac{c'' + k - \sqrt{(c''+k)^2 - \left(\frac{\|\mathbf{u}''\|^2}{f''^2} + 1\right)(c''^2 + 2kc'' + k^2 - 1)}}{\frac{\|\mathbf{u}''\|^2}{f''^2} + 1} \begin{pmatrix} -\frac{u''}{f''} \\ \frac{v''}{f''} \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ c'' \end{pmatrix} \right], \tag{5.26}$$

where $c'' = \frac{d''}{a''}$ is a new mirror parameter, which is the ratio of the d'' (distance between $\mathbf{F}_f$ and $\mathbf{C}$) and a'' (radius of the sphere), f'' is the focal length of the perspective camera

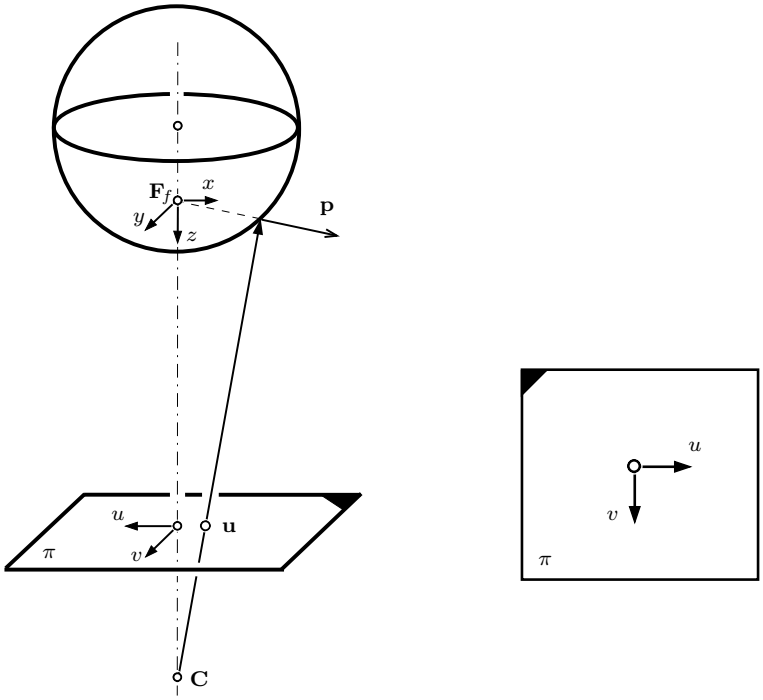


Figure 5.17.: Left: Coordinate system of the spherical catadioptric camera. The origin is located in $\mathbf{F}_f$. Right: A coordinate system in the digital image.

Figure 5.18.: A spherical catadioptric camera consisting of a perspective camera and a spherical mirror.

and $\mathbf{u}'' = (-u'', v'')^\top$ is an image point measured in the sensor Cartesian coordinate system.

We assume that the whole spherical mirror, i.e., the view field ellipse, is visible in the image. It is a consequence of the perspective projection, see Figure 5.19, that

$$\frac{r''_{max}}{f''} = \frac{a''}{\sqrt{(d'' + k\,a'')^2 - a''^2}}$$

and thus

$$f'' = r''_{max}\sqrt{\left(\frac{d''}{a''} + k\right)^2 - 1} = r''_{max}\sqrt{(c'' + k)^2 - 1}\,. \tag{5.27}$$

Equation (5.27) can be substituted into Equation (5.26) to obtain the 3D vector $\mathbf{p}''$ as a *two*-parametric function of parameters c'' and k (the position of the fictive center $\mathbf{F}_f$).

Let us investigate the relationship of the distance k of the fictive center $\mathbf{F}_f$ and the ratio $\frac{d''}{a''}$. We minimized (over k) the sum of squares of angular differences between real 3D rays reflected by a mirror surface to scene points and the approximated rays emanating from the fictive center $\mathbf{F}_f$, see Figure 5.20. It can be seen that the distance of the fictive center is changed moderately. A small change of k does not affect significantly the error arised by approximating the non-central model by the central one. Therefore, we fixed $k = 0.64$. By that we reduced the number of parameters to *one*, i.e., to $c'' = \frac{d''}{a''}$.

Result 3 *Let the functions $m(\rho, \mathbf{a}'') = c''$ and $i(\rho) = \frac{a''}{\rho}$. Then, the Theorem 1 holds for the SCD cameras and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} \simeq \frac{c + k - \sqrt{(c+k)^2 - \left(\frac{\|\mathbf{u}\|^2}{f^2} + 1\right)(c^2 + 2k\,c + k^2 - 1)}}{\frac{\|\mathbf{u}\|^2}{f^2} + 1} \begin{pmatrix} -\frac{u}{f} \\ \frac{v}{f} \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}, \tag{5.28}$$

where

$$f = r_{max}\sqrt{(c + k)^2 - 1}\,. \tag{5.29}$$

Proof. Substituting $\mathbf{u}'' = \rho\,\mathtt{R}\mathbf{u}$ into the Equation (5.26) leads to

$$\mathbf{p}'' = a''\begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix}\left[\frac{c'' + k - \sqrt{(c''+k)^2 - \left(\frac{\|\mathtt{R}\mathbf{u}\|^2}{\frac{f''^2}{\rho^2}} + 1\right)(c''^2 + 2k\,c'' + k^2 - 1)}}{\frac{\|\mathtt{R}\mathbf{u}\|^2}{\frac{f''^2}{\rho^2}} + 1}\begin{pmatrix} \frac{\mathtt{R}\mathbf{u}}{\frac{f''}{\rho}} \\ -1 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ c'' \end{pmatrix}\right] =$$

$$= a''\begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix}\left[\frac{c + k - \sqrt{(c+k)^2 - \left(\frac{\|\mathbf{u}\|^2}{f^2} + 1\right)(c^2 + 2k\,c + k^2 - 1)}}{\frac{\|\mathbf{u}\|^2}{f^2} + 1}\begin{pmatrix} \frac{\mathbf{u}}{f} \\ -1 \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ c \end{pmatrix}\right] = a''\begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix}\mathbf{p}\,,$$

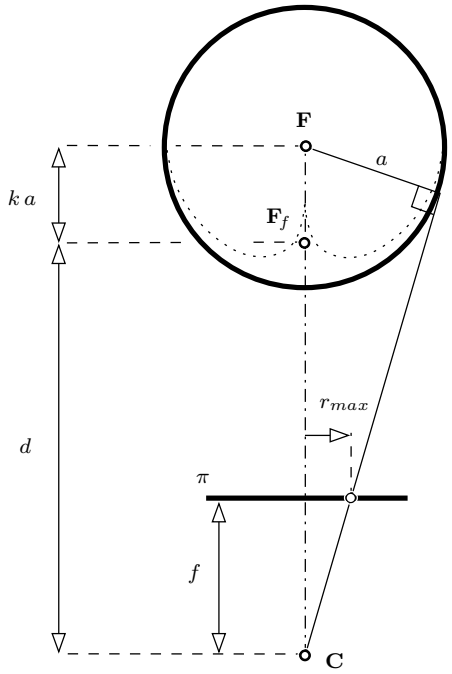


Figure 5.19.: An approximation of a non-central spherical model by a central one with a fictive viewpoint $\mathbf{F}_f$.

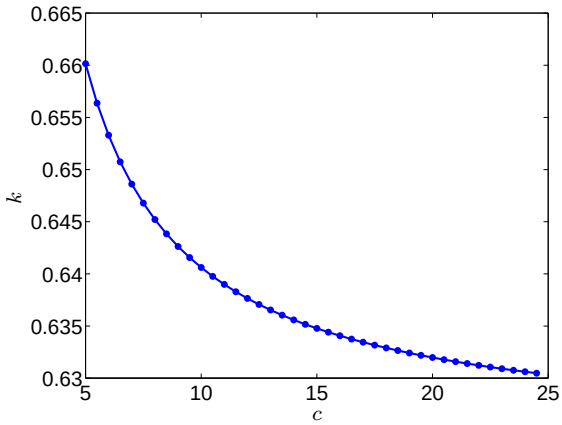


Figure 5.20.: The position k of the fictive center $\mathbf{F}_f$ as a function of the ratio $c = \frac{d}{a}$.

where

$$f \;=\; \frac{f''}{\rho} = \frac{r''_{max}}{\rho}\sqrt{(c''+k)^2-1} = r_{max}\sqrt{(c+k)^2-1}\,. \qquad \square$$

Similarly as for the HCD camera, the unknown scale ρ is absorbed into the radius r_{max} of the view field circle and $c'' = c$, since c'' is unitless variable. It follows from the above proof that the Consequence 3 in Section 5.2.1 holds for approximate central SCD cameras as well.

5.4.2. Camera auto-calibration

The auto-calibration method from epipolar geometry is analogical to the method for HCD cameras as described in Section 5.3.2. The difference is that the vector $\mathbf{p}$ from Equation (5.28) is linearized instead of the vector from Equation (5.19). We used MAPLE [49] again to obtain analytical forms of $\frac{\partial p_i}{\partial c}$ in Equation (5.21). Their forms are not listed here due to their extensiveness, see appendix A.3.

The linearization was done at the point $c_0 = 10$, which is a typical ratio of the camera distance to the radius of the sphere that can be focused. The initial estimate $c_0 \approx 10$ was sufficient in all our experiments. The original vector $\mathbf{p}$ and its linearization is shown in Figure 5.21.

There is an advantage of using SCD over HCD cameras that the ratio height/radius of the mirror does not have to be known to obtain a 9-point RANSAC. The reason is that spherical surfaces are described by a one parameter which is related to the radius of the view field circle and the camera focal length. However, the linearization of the 3D vector has to be done to get the PEP (the same as for the HCD camera), what is a disadvantage over PCD cameras where no linearization was needed.

5.4.3. Experiment

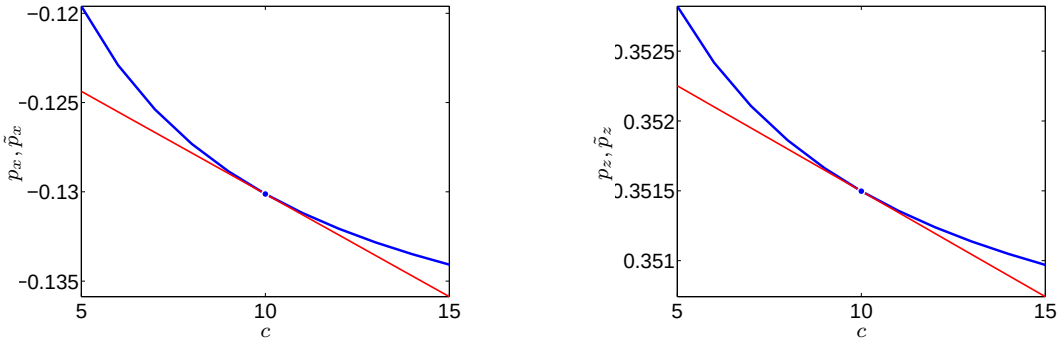
We acquired two images by a SCD camera composed of the Canon PowerShot G2 digital camera with the diameter of the view field circle equal 1042 pixels mounted above a gear ball, Figure 5.18.

The tentative correspondences, i.e., the centers of gravity of feature regions, were obtained by [58]. As a result of applying the auto-calibration method, an initial camera model and the essential matrix were obtained and outliers were rejected, see Figure 5.22.

The selected inliers and the initial motion estimate were further used in a non-linear minimization method using a precise non-central SCD model to build a 3D metric reconstruction. More details and results will be given in Section 6.3.3.

5.5. Fish-eye lens

Recently, a number of high quality, cheap, and widely available lenses with angle of view larger that 180° appeared, e.g., the Nikon FC–E8 fish-eye converter for the Nikon



(a) (b)

Figure 5.21.: Plot of the first (a) and the third (b) coordinate of the vector $\mathbf{p}$ from Equation (5.28) and its linearization $\tilde{\mathbf{p}}$ in $c_0 = 10$ for image point $\mathbf{u} = (100,\ 0)^\top$. Blue thick curves are original functions, red thin lines are the linear approximations. Blue dots represent c_0, i.e., the points of linearization.

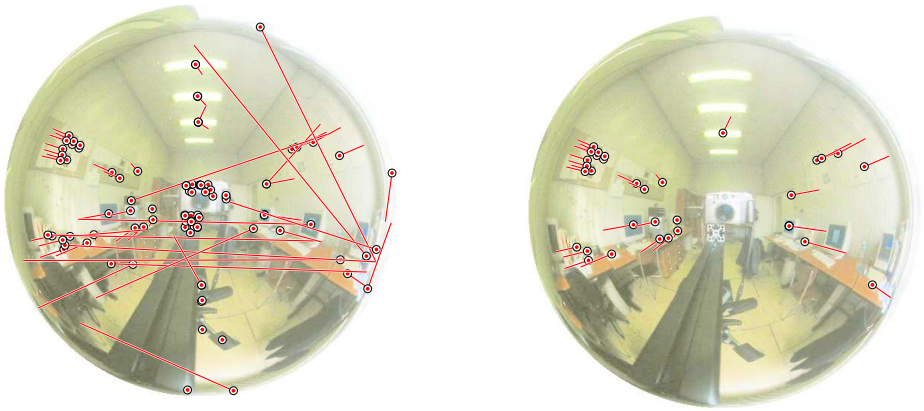


Figure 5.22.: Outliers detection during the calibration process for the SCD camera. Left: An acquired digital image from a stereopair with all tentative correspondences found by [58]. Circles mark points in the first image, lines join their matches in the next one. Right: The same image with the validated inliers only.

COOLPIX digital camera or the Sigma 8mm-f4-EX fish-eye lens for cameras with 35mm film format. The fish-eye lenses with so large field of view can be regarded as dioptric central omnidirectional cameras, see Figure 5.24.

Depending on desired accuracy, the models with various number of parameters can be derived. Here, we derive three models for both fish-eye lenses. It allows to create a hierarchy of models from the simplest (but not very accurate) to the more complicated (accurate) one. The simplest model is not very precise, however it is precise enough to reject many bad outliers. The more complicated model is more precise but it is more computational complex and it needs more points to be estimated.

5.5.1. One-parametric linear model

The linear model

$$\theta = a'' \, \|\mathbf{u}''\| \tag{5.30}$$

holds approximately true for fish-eye lenses with equi-angular projection [31, 56, 76], where a'' is a parameter, θ is the angle between a ray and the optical axis, see Figure 5.24(b). The relationship between the 3D vector $\mathbf{p}''$ emanating from the optical center $\mathbf{C}$ towards a scene point and the corresponding sensor plane point $\mathbf{u}''$, see Figure 5.23, can be expressed, according to Equation (4.3), in the lens Cartesian coordinate system as follows

$$\mathbf{p}'' = \begin{pmatrix} \mathbf{u}'' \\ g(\|\mathbf{u}''\|, a'') \end{pmatrix} = \begin{pmatrix} \mathbf{u}'' \\ \frac{\|\mathbf{u}''\|}{\tan\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{u}'' \\ \frac{\|\mathbf{u}''\|}{\tan(a'' \, \|\mathbf{u}''\|)} \end{pmatrix}. \tag{5.31}$$

Result 4 *Let the functions $m(\rho, \mathbf{a}'') = \rho \, a''$, and $i(\rho) = 1$. Then, the Theorem 1 holds for the equi-angular fish-eye lenses with one-parametric linear model and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} \simeq \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan(a \, \|\mathbf{u}\|)} \end{pmatrix}. \tag{5.32}$$

Proof. Substituting $\mathbf{u}'' = \rho \, \mathtt{R}\mathbf{u}$ into the Equation (5.31) get

$$\mathbf{p}'' = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan(\rho \, a'' \, \|\mathbf{u}\|)} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan(a \, \|\mathbf{u}\|)} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}.$$

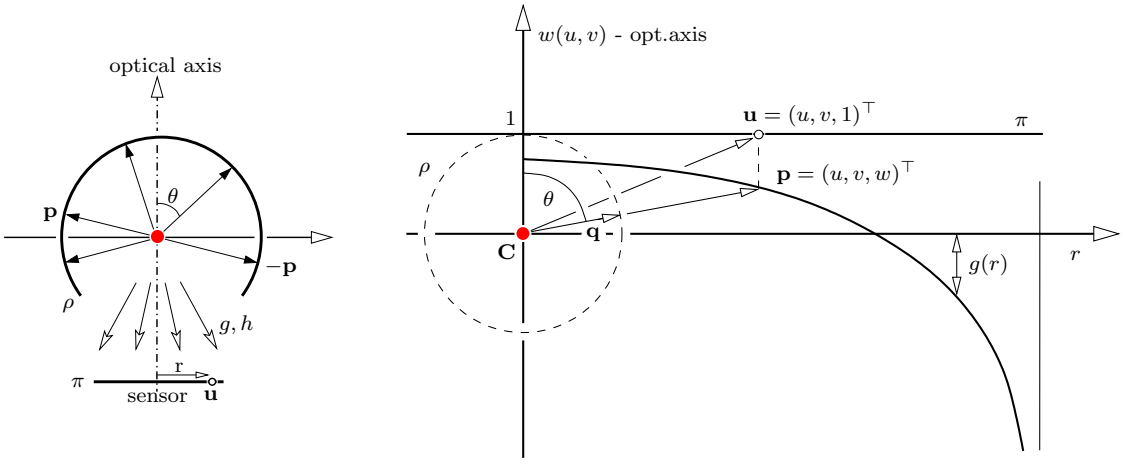
□

The unknown scale ρ is absorbed into the parameter a.

5.5.2. Two-parametric non-linear model

For the Nikon and the Sigma fish-eye lenses, respectively, we sometimes need to use more precise two-parametric models

$$\theta = \frac{a'' \, \|\mathbf{u}''\|}{1 + b'' \, \|\mathbf{u}''\|^2}, \qquad \text{resp.} \qquad \theta = \frac{1}{b''} \arcsin\left(\frac{b'' \, \|\mathbf{u}''\|}{a''}\right), \tag{5.33}$$



(a) (b)

Figure 5.23.: The geometrical interpretation of fish-eye lens projection. (a) The image point on the planar sensor π can be represented by intersecting a spherical retina ρ with camera half-rays. (b) Projection of the vector $\mathbf{q} \rightarrow \mathbf{p} \rightarrow \mathbf{u}$ into a sensor plane π through a function $g(r)$.

(a) (b)

Figure 5.24.: (a) The Nikon FC–E8 fish-eye converter. (b) All rays emanate from the projection center shown as a red bold dot.

where both a'' and b'' are parameters of the models. In general, the models may have various forms determined by the lens design and by desired accuracy.

The both models in Equation (5.33) were obtained experimentally in the following manner. We knew that the Nikon and the Sigma lenses were possessing equi-angular projection. Therefore, we took the equi-angular model as the base and added an extra parameter b'' to model the lenses more precisely. In this case, the parameter a'' captures the equi-angular projection. The parameter b'' captures other distortions since the lenses often do not satisfy equi-angular projection absolutely precisely. If $b'' = 0$ the non-linear models, Equation (5.33), become linear, Equation (5.30). Adding an extra parameter b'' is often tricky. Various more-parametric models can be found in [31, 56, 4, 76]. We tried some combinations of known models and finally we ended up with those described by Equation (5.33) as a compromise between the accuracy and the number of parameters.

The vector $\mathbf{p}''$ for the Nikon and the Sigma fish-eye lens, respectively, becomes

$$\mathbf{p}'' = \begin{pmatrix} \mathbf{u}'' \\ \frac{\|\mathbf{u}''\|}{\tan\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{u}'' \\ \frac{\|\mathbf{u}''\|}{\tan\frac{a''\,\|\mathbf{u}''\|}{1+b''\,\|\mathbf{u}''\|^2}} \end{pmatrix} \qquad \text{resp.} \qquad \begin{pmatrix} \mathbf{u}'' \\ \frac{\|\mathbf{u}''\|}{\tan\left(\frac{1}{b''}\arcsin\left(\frac{b''\|\mathbf{u}\|}{a''}\right)\right)} \end{pmatrix}. \tag{5.34}$$

Result 5 *Let the functions* $m(\rho, \mathbf{a}'') = \begin{pmatrix} \rho\, a'' \\ \rho^2\, b'' \end{pmatrix}$ *and* $i(\rho) = 1$. *Then, the Theorem 1 holds for the Nikon fish-eye lens with two-parametric model and the vector* $\mathbf{p}$ *is as follows*

$$\mathbf{p} \simeq \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\frac{a\,\|\mathbf{u}\|}{1+b\,\|\mathbf{u}\|^2}} \end{pmatrix}. \tag{5.35}$$

Proof.

$$\mathbf{p}'' = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\frac{\rho\,a''\,\|\mathbf{u}\|}{1+\rho^2\,b''\,\|\mathbf{u}\|^2}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\frac{a\,\|\mathbf{u}\|}{1+b\,\|\mathbf{u}\|^2}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}\,. \qquad \square$$

The unknown scale ρ is absorbed into both parameters.

Result 6 *Let the functions* $m(\rho, \mathbf{a}'') = \begin{pmatrix} \frac{a''}{\rho} \\ b'' \end{pmatrix}$ *and* $i(\rho) = 1$. *Then, the Theorem 1 holds for the Sigma fish-eye lens with two-parametric model and the vector* $\mathbf{p}$ *is as follows*

$$\mathbf{p} \simeq \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\left(\frac{1}{b}\arcsin\left(\frac{b\|\mathbf{u}\|}{a}\right)\right)} \end{pmatrix}. \tag{5.36}$$

Proof.

$$\mathbf{p}'' = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\left(\frac{1}{b''}\arcsin\left(\frac{b''\|\mathbf{u}\|}{\frac{a''}{\rho}}\right)\right)} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan\left(\frac{1}{b}\arcsin\left(\frac{b\|\mathbf{u}\|}{a}\right)\right)} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p}\,. \qquad \square$$

The unknown scale ρ is absorbed into the parameter a only. The parameter b'' remains unchanged since it is a unitless variable.

5.5.3. One-parametric non-linear model

Suppose that the maximal view angle θ_{max} of the fish-eye lens is known (given by a manufacturer). The maximal radius r''_{max} corresponding to the θ_{max} can be easily obtained from the view field circle in the pre-calibrated image, see Figure 5.13. It allows to express the parameter a'' using the radius r''_{max} and the angle θ_{max}. From models in Equation (5.33) for the Nikon and the Sigma fish-eye lens, respectively, we obtain

$$a'' = \frac{(1 + b'' r''^2_{max})\theta_{max}}{r''_{max}} , \qquad \text{resp.} \qquad a'' = \frac{b''\, r''_{max}}{\sin(b''\, \theta_{max})} . \tag{5.37}$$

Substituting Equation (5.37) into Equation (5.33) yields a *one*-parametric nonlinear model

$$\theta = \frac{(1 + b'' r''^2_{max})\theta_{max}\, \|\mathbf{u}''\|}{r''_{max}(1 + b''\|\mathbf{u}''\|^2)} , \qquad \text{resp.} \qquad \theta = \frac{1}{b''} \arcsin \frac{\|\mathbf{u}''\| \sin(b''\theta_{max})}{r''_{max}} . \tag{5.38}$$

Result 7 *Let the functions $m(\rho, \mathbf{a}'') = \rho^2\, b''$ and $i(\rho) = 1$. Then, the Theorem 1 holds for the Nikon fish-eye lens with one-parametric non-linear model and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} \simeq \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan \frac{(1+br^2_{max})\theta_{max}\, \|\mathbf{u}''\|}{r_{max}(1+b\|\mathbf{u}\|^2)}} \end{pmatrix} . \tag{5.39}$$

Proof.

$$\mathbf{p}'' = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan \frac{(1+\rho^2\, b''\, \frac{r''^2_{max}}{\rho^2})\theta_{max}\, \|\mathbf{u}\|}{\frac{r''_{max}}{\rho}(1+\rho^2\, b''\|\mathbf{u}\|^2)}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\tan \frac{(1+b\, r^2_{max})\theta_{max}\, \|\mathbf{u}\|}{r_{max}(1+b\, \|\mathbf{u}\|^2)}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p} . \qquad \square$$

The unknown scale ρ is absorbed into the parameter b and in the radius r_{max}.

Result 8 *Let the functions $m(\rho, \mathbf{a}'') = b''$ and $i(\rho) = 1$. Then, the Theorem 1 holds for the Sigma fish-eye lens with one-parametric non-linear model and the vector $\mathbf{p}$ is as follows*

$$\mathbf{p} \simeq \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\frac{1}{b} \arcsin \frac{\|\mathbf{u}\| \sin(b\, \theta_{max})}{r_{max}}} \end{pmatrix} . \tag{5.40}$$

Proof.

$$\mathbf{p}'' = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\frac{1}{b''} \arcsin \frac{\|\mathbf{u}\| \sin(b''\, \theta_{max})}{\frac{r''_{max}}{\rho}}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \frac{\|\mathbf{u}\|}{\frac{1}{b} \arcsin \frac{\|\mathbf{u}\| \sin(b\, \theta_{max})}{r_{max}}} \end{pmatrix} = \rho \begin{pmatrix} \mathtt{R} & \\ & 1 \end{pmatrix} \mathbf{p} . \qquad \square$$

The unknown scale ρ is absorbed into the radius r_{max}. The parameter b'' remains unchanged since it is a unitless variable.

All the results formulated for the Nikon and the Sigma fish-eye lenses have shown the absorption of the ρ into parameters. Therefore, the Consequence 3 in Section 5.2.1 holds for both fish-eye lenses with all models presented here.

5.5.4. Camera auto-calibration

The three above models allow to do the auto-calibration in three steps:

1. one-parametric linear model,
2. one-parametric non-linear model,
3. two-parametric non-linear model,

such that we begin with the simplest model and end up with the most complex one.

The estimation hierarchy allows to gradually reject outliers depending on their error. The basic idea is that the mismatches with large error do not fulfil neither the most precise model nor the less precise one. The worst outliers can thus be detected by a simpler model and a more complicated two-parametric model can be estimated from data with fewer number of outliers. Using the more accurate (but more computational expensive) model on the inliers found by the simpler models dramatically speeds up the convergence of the RANSAC-based algorithm. However, the point sampling in the RANSAC is done from the inliers found in the previous stage but the error is always computed on all correspondences. This idea is employed in the LO-RANSAC technique explained in more detail in [18].

In the next we will describe the estimation process for the Nikon fish-eye lens. For the Sigma fish-eye lens the procedure is analogical.

Auto-calibration for the one-parametric linear model

The same procedure, as for catadioptric cameras, can be applied to fish-eye lenses to get the PEP. Since $h(\|\mathbf{u}\|) = 1$ in fish-eye models, see Equation (4.3), only a linearization of the third coordinate of $\mathbf{p} = \left(\mathbf{u}^\top,\ g(\|\mathbf{u}\|, a)\right)^\top$, Equation (5.31), will be necessary. The linear part $\tilde{g}(r, a)$ of Taylor series of function $g(.)$ in the point a_0 is the following

$$\tilde{g}(r, a) = g(r, a_0) + g_a(r, a_0)(a - a_0)\,, \tag{5.41}$$

where $r = \|\mathbf{u}\|$, and

$$\begin{aligned} g(r, a_0) &= \frac{r}{\tan(a_0\, r)}\,, \\ g_a(r, a_0) &= -\left(1 + (\tan(a_0\, r))^2\right)\left(\frac{r}{\tan(a_0\, r)}\right)^2. \end{aligned} \tag{5.42}$$

Function $g_a(r, a_0)$ is the partial derivative of $g(r, a)$ w.r.t. a evaluated in a_0. The vector

$$\tilde{\mathbf{p}} = \begin{pmatrix} \mathbf{u} \\ g(.) - a_0\, g_a(.) + a\, g_a(.) \end{pmatrix} = \begin{pmatrix} \mathbf{u} \\ w \end{pmatrix} + a \begin{pmatrix} \mathbf{0} \\ s \end{pmatrix} = \mathbf{x} + a\,\mathbf{s} \tag{5.43}$$

is obtained by linearizing Equation (5.32). Vectors $\mathbf{x}$ and $\mathbf{s}$ are computed from image coordinates of the corresponding points. The epipolar constraint, Equation (4.20), can be used for linearized vectors, i.e., $\tilde{\mathbf{p}}_2^\top \mathtt{F}\, \tilde{\mathbf{p}}_1 = 0$, and arranged to the PEP

$$(\mathtt{D_1} + a\,\mathtt{D_2} + a^2\mathtt{D_3})\mathbf{f} = \mathbf{0}\,. \tag{5.44}$$

We end up with the same equation as for HCD cameras described in Section 5.3.2, just matrices $\mathtt{D}_i$ are composed differently,

$$\begin{aligned}
\mathtt{D_1} &= \begin{bmatrix} u_2u_1 & u_2v & u_2w_1 & v_2u_1 & v_2v_1 & v_2w_1 & w_2u_1 & w_2v_1 & w_2w_1 \end{bmatrix}, \\
\mathtt{D_2} &= \begin{bmatrix} 0 & 0 & u_2s_1 & 0 & 0 & v_2s_1 & s_2u_1 & s_2v_1 & w_2s_1 + s_2w_1 \end{bmatrix}, \\
\mathtt{D_3} &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_2s_1 \end{bmatrix}.
\end{aligned} \tag{5.45}$$

The vector $\mathbf{f}$ has the form as in Equation (5.12).

The method leads to a 9-point RANSAC since at least 9 points are needed to solve the PEP in Equation (5.44).

Auto-calibration for the one-parametric non-linear model

The same procedure as for the one-parametric linear model can be done. Suppose that r_{max} and θ_{max} be known. The linearization of Equation (5.39) w.r.t. b yields $\tilde{\mathbf{p}}$ in Equation (5.43), where

$$\begin{aligned}
g(r, b_0) &= \frac{r}{\tan\left(\frac{(1+b_0\, r_{max}^2)\theta_{max} r}{r_{max}(1+b_0\, r^2)}\right)}, \\
g_b(r, b_0) &= -\frac{\theta_{max} r^2 (1 + g(r, b_0)^2)(r_{max}^2 - r^2)}{g(r, b_0)^2 (1 + b_0 r^2)^2 r_{max}}\,.
\end{aligned} \tag{5.46}$$

The vectors $\tilde{\mathbf{p}}_1$ and $\tilde{\mathbf{p}}_2$ using Equation (5.43) with Equation (5.46) and Equation (4.20) yields the the same PEP as in Equation (5.44). Matrices $\mathtt{D}_i$ contain the same elements as in Equation (5.45).

The one-parametric non-linear model leads to a 9-point RANSAC again. This model is more precise than the one-parametric linear model since it captures the non-linearity between the angle θ and the radius r.

Auto-calibration for the two-parametric non-linear model

The models in Equation (5.33) are two-parametric non-linear functions. To arrive at the PEP the function $g(.)$ in the vector $\mathbf{p}$ in Equation (5.35) has to be linearized w.r.t. a and b in a_0 and b_0.

$$\tilde{g}(r, a, b) = g(r, a_0, b_0) + g_a(r, a_0, b_0)(a - a_0) + g_b(r, a_0, b_0)(b - b_0)\,, \tag{5.47}$$

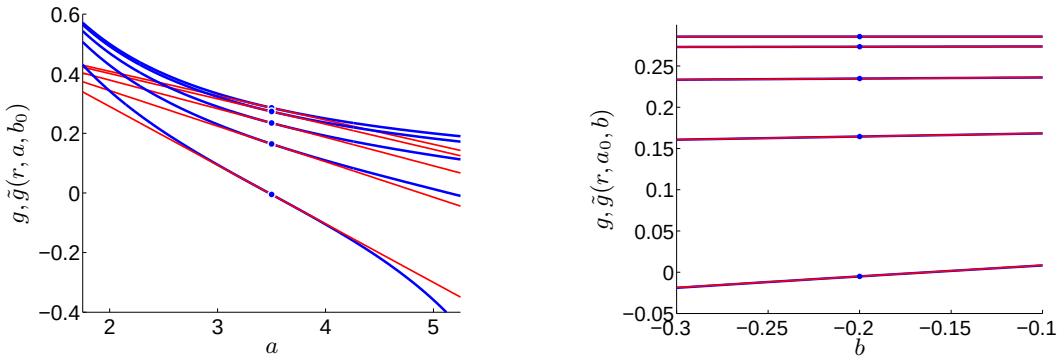


Figure 5.25.: Comparison of the graph of original function g (thick blue curves) and its linearization $\tilde{g}$ (red lines). Left: Change of the parameter a for fixed $b = -0.2$. Right: Change of the b for the fixed $a = 3.5$. Graphs are for radii $r = 10$ (upper), $100, 200, 300, 435$pxl (lower), where $r_{max} = 435$pxl $\sim 91.5°$. Blue dots mean points of the Taylor expansion: $a_0 = 3.5$, $b_0 = -0.2$. Parameters a, b were changed by $\pm 50\%$.

where

$$\begin{aligned} g(r, a_0, b_0) &= \frac{r}{\tan \frac{a_0 r}{1+b_0 r^2}}\,, \\ g_a(r, a_0, b_0) &= -\frac{1 + \left(\tan \frac{a_0 r}{1+b_0 r^2}\right)^2}{(1 + b_0 r^2)}\, g(r, a_0, b_0)^2\,, \\ g_b(r, a_0, b_0) &= -\frac{a\, r^2}{(1 + b_0 r^2)}\, g_a(r, a_0, b_0)\,. \end{aligned}$$

The functions $g_a(.)$, resp. $g_b(.)$ are the partial derivatives of $g(r, a, b)$ w.r.t. a, resp. b, evaluated in the point a_0, resp. b_0.

The vector $\tilde{\mathbf{p}}$ can be written using Equation (5.47) as follows

$$\begin{aligned} \tilde{\mathbf{p}} &= \begin{pmatrix} \mathbf{u} \\ g(.) - a_0 g_a(.) - b_0 g_b(.) \end{pmatrix} + a \begin{pmatrix} \mathbf{0} \\ g_a(.) \end{pmatrix} + b \begin{pmatrix} \mathbf{0} \\ g_b(.) \end{pmatrix} = \\ &= \begin{pmatrix} \mathbf{u} \\ w \end{pmatrix} + a \begin{pmatrix} \mathbf{0} \\ s \end{pmatrix} + b \begin{pmatrix} \mathbf{0} \\ t \end{pmatrix} = \\ &= \mathbf{x} + a\,\mathbf{s} + b\,\mathbf{t}\,, \end{aligned}$$

where $\mathbf{x}$, $\mathbf{s}$ and $\mathbf{t}$ are vectors computed from image coordinates of corresponding points. The epipolar constraint, Equation (4.20), can be used for linearized vectors, i.e. $\tilde{\mathbf{p}}_2^\top \mathtt{F}\, \tilde{\mathbf{p}}_1 = 0$, as

$$(\mathbf{x}_2 + a\,\mathbf{s}_2 + b\,\mathbf{t}_2)^\top \mathtt{F}\, (\mathbf{x}_1 + a\,\mathbf{s}_1 + b\,\mathbf{t}_1) = 0\,, \tag{5.48}$$

and arranged to the PEP

$$(\mathtt{D}_1 + a\,\mathtt{D}_2 + a^2\mathtt{D}_3)\,\mathbf{l} = \mathbf{0}\,. \tag{5.49}$$

The matrices $\mathtt{D}_\mathtt{i}$ and the vector $\mathbf{l}$ are as follows

$$\begin{aligned}
\mathtt{D}_1 &= \big[\; u_1u_2 \;\; v_1u_2 \;\; w_1u_2 \;\; u_1v_2 \;\; v_1v_2 \;\; w_1v_2 \;\; u_1w_2 \;\; v_1w_2 \;\; w_1w_2 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t_1u_2 \;\; t_1v_2 \;\; u_1t_2 \;\; v_1t_2 \;\; t_1w_2 + w_1t_2 \;\; t_1t_2 \;\big], \\
\mathtt{D}_2 &= \big[\; 0 \;\; 0 \;\; s_1u_2 \;\; 0 \;\; 0 \;\; s_1v_2 \;\; u_1s_2 \;\; v_1s_2 \;\; s_1w_2 + w_1s_2 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; t_1s_2 + s_1t_2 \;\; 0 \;\big], \\
\mathtt{D}_3 &= \big[\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; s_1s_2 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\big], \\
\mathbf{l} &= (\; F_{11} \;\; F_{12} \;\; F_{13} \;\; F_{21} \;\; \ldots \;\; F_{33} \;\; b\,F_{13} \;\; b\,F_{23} \;\; b\,F_{31} \;\; b\,F_{32} \;\; b\,F_{33} \;\; b^2F_{33} \;)^\top .
\end{aligned} \tag{5.50}$$

The vector $\mathbf{l}$ contains elements of the fundamental matrix and additional six (dependent) products $b\,\mathtt{F}_{ij}$. The parameter b can be determined from any of them. The matrices $\mathtt{D}_i$ should be square and therefore at least 15 points matches are needed for solving the PEP in Equation (5.49). There are 30 solutions of a, many of them are zero, infinite or complex. For every a, six possible solutions of b exist. Since there is noise in data, we choose the solution which best fits our model in terms of the smallest error given by Oliensis, Equation (4.22). As in the previous sections, the least-square solution by left-multiplication of Equation (5.49) by $\mathtt{D}_1^\top$ can be used with more than 15 points.

The functions $\tilde{g}$ in Equation (5.47) and g in Equation (5.34) are shown in Figure 5.25. As can be seen, $b_0 = 0$ can be chosen thanks to the linear character of $g(r, a_0, b)$. We assume that $a_0 = \frac{\theta_{max}}{r_{max}}$ is known precisely enough to be used as a point of Taylor series expansion.

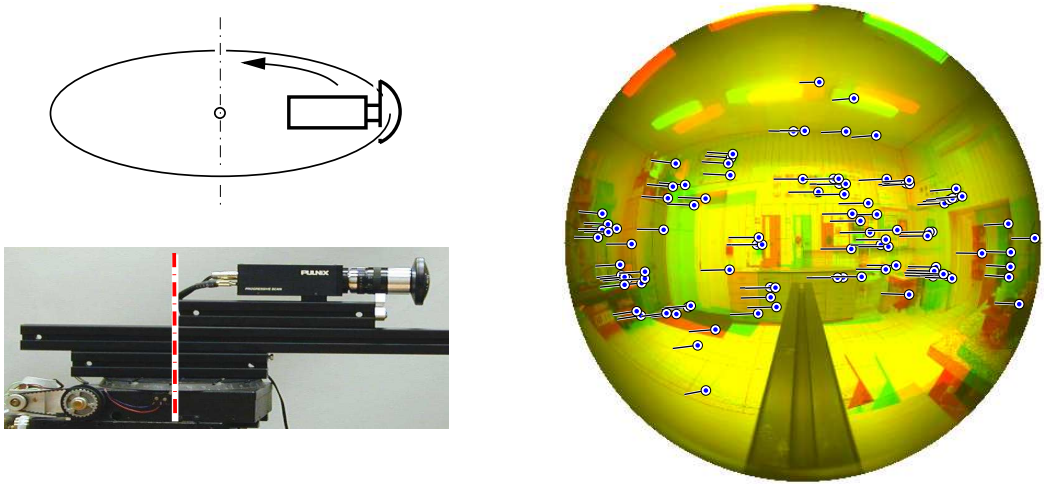
5.5.5. Experiments

Trajectory estimation

We carried out the same circular experiment as for the PCD and the HCD camera, described in Section 5.2.3 and Section 5.3.3. The PULNIX digital camera with the Nikon FC-E8 fish-eye lens were used. The diameter of the view field circle was 870 pixels. Images were acquired continuously, however, only images in 10° step were selected, 36 images in total. Correspondences were obtained by tracking in the commercial program boujou [48]. The data was provided by Josef Šivic [79].

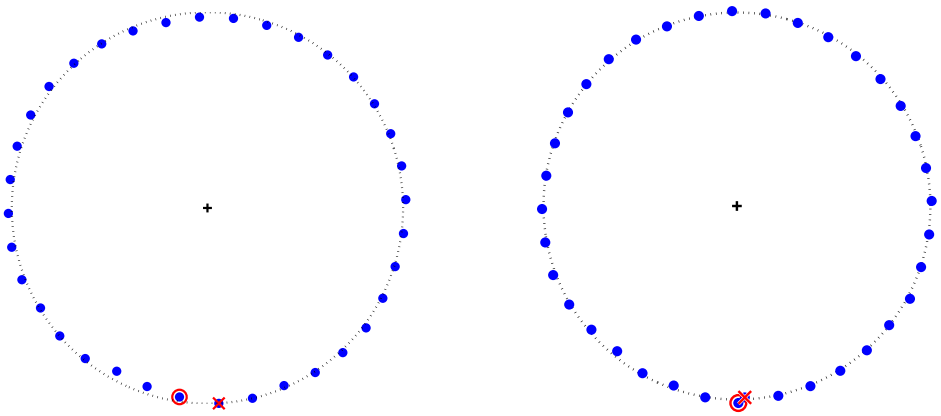
The same procedure as for the PCD and the HCD camera was carried out, see Section 5.2.3 for more details. First, the parameters a, b, and the essential matrix $\mathtt{F}$ were computed for every consecutive pair in the sequence, see Figure 5.27(a). Second, one $\hat{a}$ and one $\hat{b}$ were computed as the mean over all computed a's and b's and used to re-compute the essential matrices again for the whole sequence, see Figure 5.8(b).

The calibration procedure was done in three steps. First, a 9-point RANSAC with the linear model, Equation (5.30), and then, a 9-point RANSAC with the non-linear model, Equation (5.38), as a pre-test to detect most of outliers were used. Finally, a 15-point



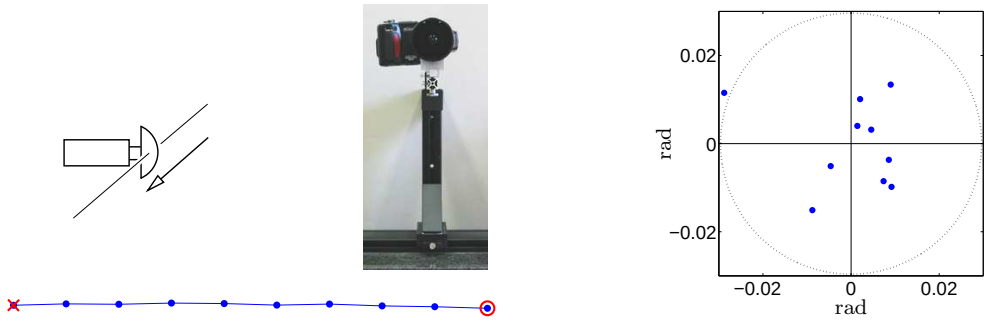
(a) (b)

Figure 5.26.: The circular sequence. (a) The Nikon FC–E8 fish-eye converter mounted on the PULNIX TM1001 digital camera is rotated along a circle. Data acquisition by courtesy of J.Šivic [79]. (b) Correspondences between two consecutive images. Circles mark points in the first image, lines join their matches in the next one. The two images are superimposed in red and green channel.



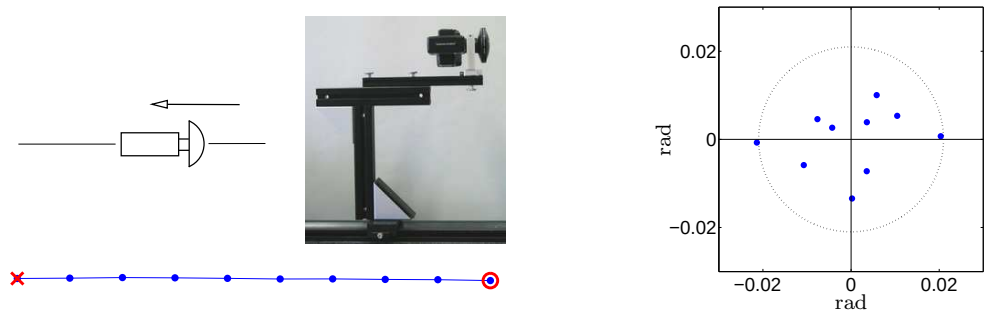
(a) (b)

Figure 5.27.: Estimated trajectories for the circular sequence. Red ∘ depicts the starting position, × depicts the end position. (a) The trajectory estimated by the calibration. Different a, b for each pair. (b) The trajectory estimated after the calibration using the same $\hat{a}$, $\hat{b}$ for each consecutive pair.



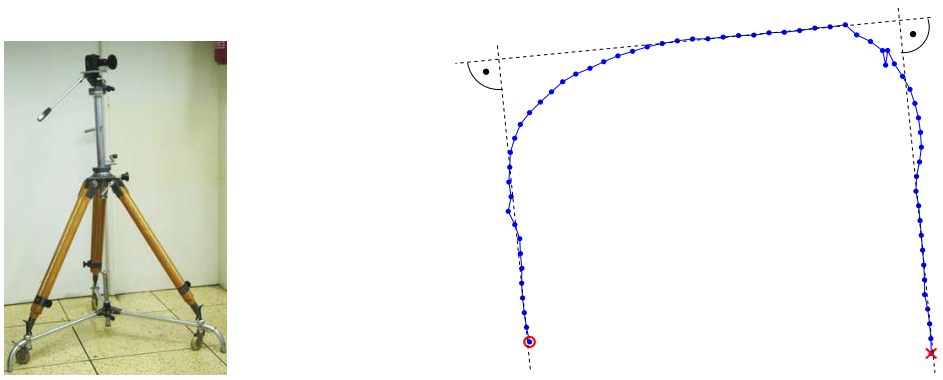
(a) (b)

Figure 5.28.: Side motion. The Nikon FC–E8 fish-eye converter with COOLPIX digital camera was used. (a) Top left: schematic diagram of the camera motion. Top right: real setup. Bottom: the estimated trajectory. (b) Angular error between the direction of motion and the optical axis for each pair, and circle 3σ is plotted.



(a) (b)

Figure 5.29.: Forward motion of the fish-eye lens. See Figure 5.28 for the explanation.



(a) (b)

Figure 5.30.: General motion. (a) Setup of the experiment. A mobile tripod with the camera. (b) The estimated trajectory.

RANSAC with the non-linear model, Equation (5.33), was run to compute parameters a, b, and $\mathtt{F}$.

It can be seen in Figure 5.27(a) that the trajectory is circular but there is a discrepancy between the starting and the end positions. The second, result in Figure 5.27(b) shows that the end position correctly approaches the starting one. However, both results are good enough as starting points for a non-linear bundle adjustment.

The next experiment calibrates the omnidirectional camera (the Nikon FC–E8 fish-eye converter with the Nikon COOLPIX digital camera with resolution 1600×1200 pxl and the diameter of the view field 1172 pixels) from its translation in the direction perpendicular to the optical axis, see Figure 5.28, and in the direction along the optical axis, see Figure 5.29. Estimated trajectories are depicted in Figure 5.28(a) and Figure 5.29(a). The angular differences between estimated and true motion directions for every pair are depicted in Figure 5.28(b) and Figure 5.29(b). In fact, the errors are approximately the same. The difference is under experimental error and therefore, in this case, the side and the forward motion perform similarly.

The next experiment shows the calibration from a general planar motion. Figure 5.30(a) shows a mobile tripod with an omnidirectional camera. Figure 5.30(b) shows an estimated trajectory. We made a U-shaped trajectory with right angles. Discontinuities of the trajectory were caused by hand driven motion of the mobile tripod. Naturally, they have no effect on the final estimate and the final trajectory has really right angles.

3D metric reconstruction

In this experiment, one image pair was selected from the Venice Yard QY dataset, acquired by the Sigma 8mm-f4-EX fish-eye lens with view angle 180° mounted on the Canon EOS–1Ds digital camera with resolution 4064×2704 pxl. The diameter of the view field circle was 2528 pixels. The tentative correspondences, i.e., centers of gravity of every region, were obtained by [58]. As a result of applying the auto-calibration method, the camera model and the essential matrix were obtained and most outliers were rejected. The same hierarchical approach as in the previous experiment (the 9-points RANSAC followed by the 15-points one) was applied. The obtained calibrated cameras and validated point matches were used for 3D reconstruction achieved by a linear technique described in Section 4.5.

Final 3D reconstruction was improved by a non-linear bundle adjustment (courtesy of Tomáš Werner) tuning all 3D points $\mathbf{X}_i$, both camera matrices $\mathtt{P}_i$ in Equation (4.24) and camera model parameters $(u_0, v_0)^\top$, a, and b. The bundler enforced the same internal parameters for both cameras. To show the quality of the 3D reconstruction, some correspondences, like corners on the walls, have been established manually. The estimated camera matrices were used for reconstructing these points. Finally, textures were mapped on the planar rectangles formed by the reconstructed 3D points. See Figure 5.31 for the result. Notice how *precise* and *complete* a 3D reconstruction from only *two* omnidirectional images can be obtained. The RMS of the reprojection error was 0.25 pxl.

Figure 5.31.: 3D metric reconstruction of a Venice yard from *two* uncalibrated omnidirectional images with automatically detected point correspondences. Top: One view of 3D reconstruction of a scene together with both cameras. Bottom: Stereo omnidirectional images acquired by the Sigma fish-eye lens. On the right image the correspondences marked by the auto-calibration method as inliers are shown.

Figure 5.32.: 3D metric reconstruction of a Venice yard from *many* uncalibrated omnidirectional images by projective factorization method with automatically detected point correspondences. Top: One view of the 3D reconstruction. Bottom Left: Result after full automatic steps. Bottom Right: The same view with manually added planar textures.

In the next experiment, all 13 images from the Venice Yard QY dataset were used. The experiment shows that the auto-calibration method can be used in a multiple view reconstruction together with a factorization technique [57]. See paper [60] for more details about the factorization method applied to omnidirectional images. The camera model was obtained by the proposed auto-calibration method. All consecutive image pairs were used to estimate camera parameters and a metric reconstruction has been made by the projective factorization method [57] followed by a metric upgrade, see [60]. Finally, the same bundle-adjustment as in the previous experiment was applied to obtain a more precise 3D metric reconstruction. See Figure 5.32 for the 3D reconstruction.

Let us emphasize that the calibration procedure described is fully automatic till the step shown in Figure 5.32 (bottom left). After this step isolated 3D points (reconstructed from automatically established and validated point correspondences) and positions of all cameras are obtained. Then, planar textures in Figure 5.32 (bottom right) have been added manually.

In the last experiment, the camera (the Nikon FC–E8 fish-eye lens mounted on the Nikon COOLPIX digital camera) was moving along a straight line with constant distance steps (15 cm) capturing the scene in the direction perpendicular to the motion. The same technique as in the previous experiment was used (the auto-calibration followed by the factorization technique and the bundle adjustment). See some of input images and the final 3D reconstruction in Figure 5.33. Notice in Figure 5.33(c) that the estimated trajectory is really straight and distances between all 11 cameras are equal. It can be seen in the top-view in Figure 5.33(b) that all reconstructed points are correctly in the field of view of the cameras.

5.6. Robust estimation based on RANSAC with bucketing

We designed all the auto-calibration methods in such a way that they lead to the PEP. The PEP allows to estimate simultaneously the camera model and essential matrix from a given minimal subset of points. As we have already mentioned, the robust technique, e.g., RANSAC [29, 43], can be used to handle correspondences with outliers. Our methods lead to either a 9- or a 15-point RANSAC. In this section we focus on problem arising in the RANSAC-based estimation technique for omnidirectional images when the most of correspondences are established near the view field center.

The automatic search for correspondences in omnidirectional images becomes more complicated than in perspective images because the affine invariance of corresponding features, used by most of methods, is preserved only approximately for large motions of the camera. The "non-affinity" of regions is more evident near the boundary of the image, see Figure 5.34, and therefore, usually, most correspondences are established near the center. Such correspondences satisfy the camera model for almost any degree of image formation non-linearity. They are often all selected in a RANSAC as inliers and the estimation stops prematurely. The most informative points near the border of the view field are not used and an incorrect camera model is estimated. We show [63]

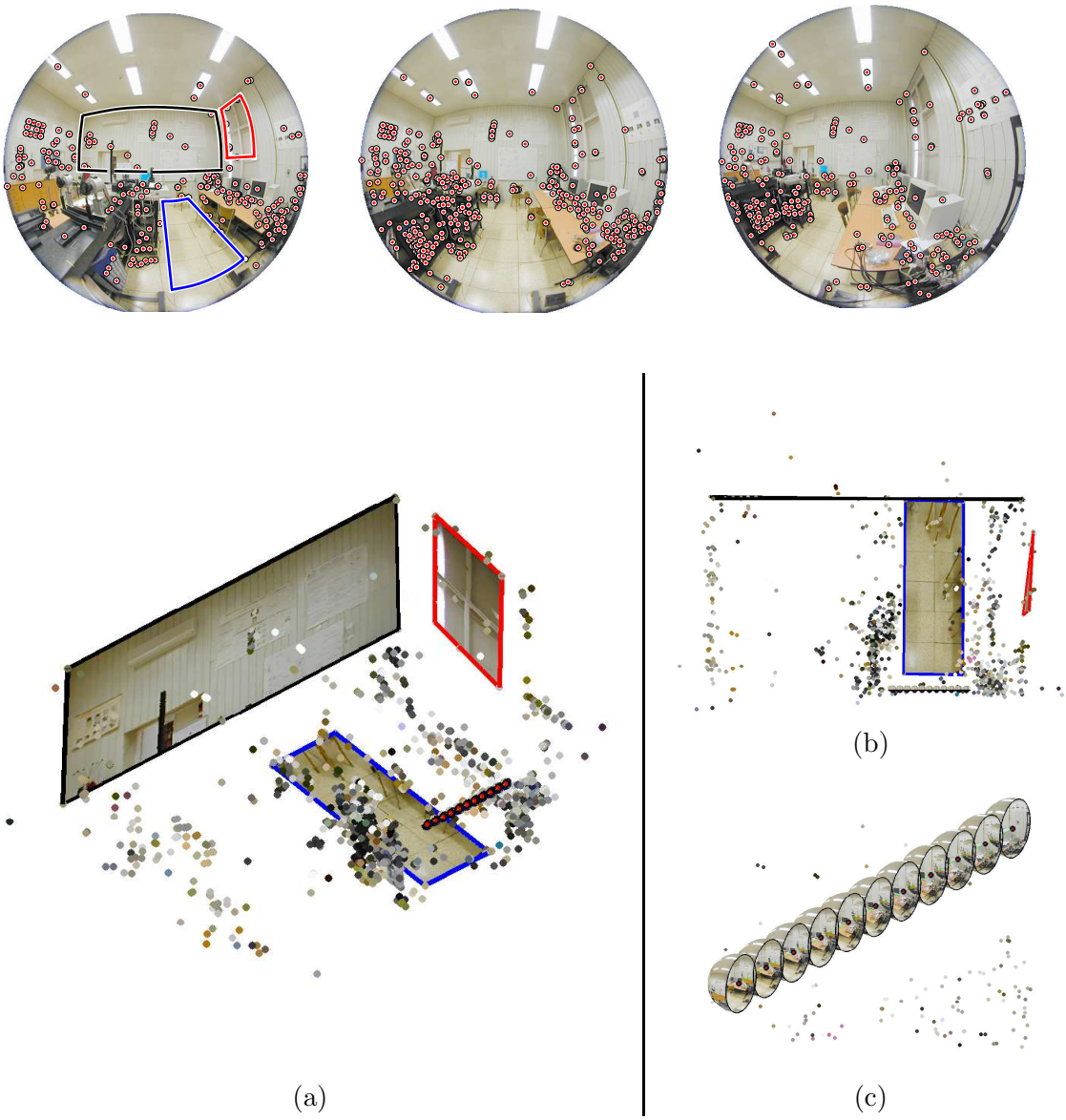
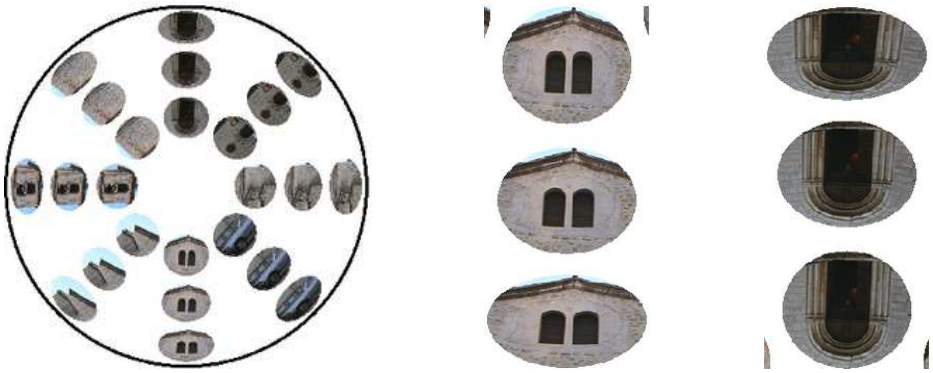


Figure 5.33.: 3D metric reconstruction of the CMP lab from *many* uncalibrated omnidirectional images by the projective factorization method. Top: Three images from a sequence, the first, the middle, and the last. (a) One view of the 3D reconstruction with three planar textures. (b) Birds' eye view of the lab. (c) Detail view of the reconstructed cameras represented by spherical retinas.



(a) (b) (c)

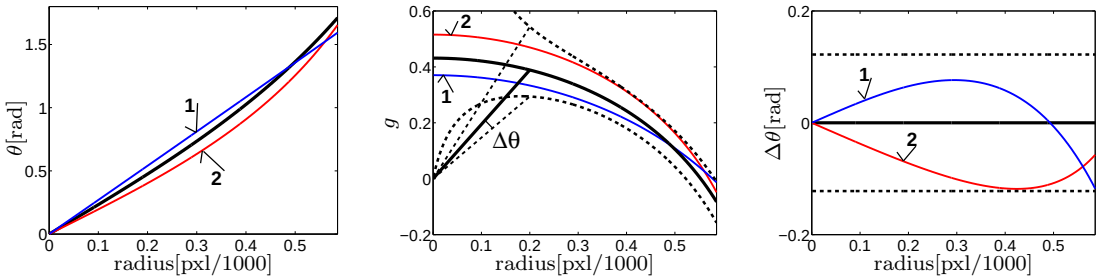
Figure 5.34.: Affine invariance of planar regions is approximately preserved in omnidirectional images. (a) The image of an object gets non-affinely distorted as it moves towards the boundary of the view field circle. The distortion is significant only for large regions and large motions. (b),(c) Detail views.

that a remedy to this problem is achieved by not using points near the center of the view field circle for the camera model estimation and controlling the points sampling in the RANSAC. We propose a strategy for the points sampling, similar to the bucketing in [100], in order to obtain a good estimate in a reasonable time.

This section is most relevant to the fish-eye lenses for which we had to add the second parameter b in the camera model to capture a non-ideal equi-angular projection. We show that the polynomial approximation, often used, is not very good. We will explain the problems on the Nikon fish-eye lens. The problem is analogical for the Sigma fish-eye lens and it leads to the same conclusion. The bucketing technique, proposed in the next, can be used for catadioptric cameras as well since it helps to speed up the RANSAC.

5.6.1. Tolerance in model fitting

As it was described earlier, an angle between a ray and its corresponding epipolar plane is used as the criterion of the estimation quality, see Equation (4.22). Ideally, it should be zero but we admit some tolerance in real situations. The tolerance in the angular error propagates into the tolerance in camera model parameters, see Figure 5.35. The region in which models satisfy a fixed tolerance $\Delta\theta$ is narrowing when the radius of points in the image is increasing, see Figure 5.35(b). Since $g(r = 0, a, b) = \frac{1}{a}$ (see appendix A.1), the points near the center $(u_0, v_0)^\top$ will affect only the parameter a. There is large tolerance in the parameter a because the tolerance region near the center $(u_0, v_0)^\top$ is large. If the majority of points is near the center then RANSAC finds the model with high number of inliers there and stops prematurely, see Figure 5.36(a), because the RANSAC is stopped if the probability of finding more inliers [43] drops below a defined threshold, usually 5%. However, there often exists a model with more inliers that suits to the points



(a) (b) (c)

Figure 5.35.: Model fitting with tolerance $\Delta\theta$. (a) The graph $\theta = k(r)$ for ground truth data (black thick curve) and two models satisfying the tolerance (red and blue curves). Parameters a and b can vary for models satisfying the tolerance. (b) The area between dashed curves is determined by the error. In this area, all models satisfying the tolerance must lie. (c) The angular error for both models w.r.t. the ground truth model.

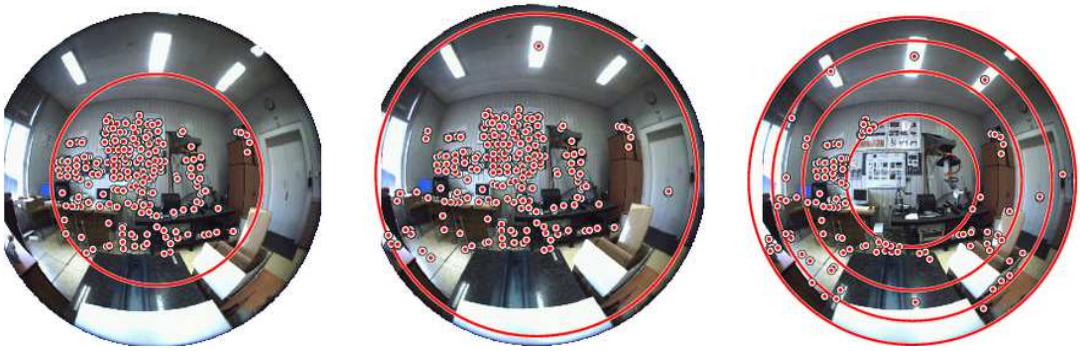
near the center as well as to the points near to the boundary of the view field circle, see Figure 5.36(b). Figure 5.37 shows how the estimation of the camera model parameter is improved (the standard deviation is smaller) by controlling the point sampling in the RANSAC. Previous analysis indicates that the points near the center of the view field circle do not sufficiently capture the correct camera model parameters.

5.6.2. Prediction error

The principle of the RANSAC estimation technique is to fit a model on a minimal subsets of points, verify the model on all data, and chose the model which supports the highest number of points [29, 43]. The question is how to select the minimal subset to obtain the best estimate.

We are going to investigate how a point selection can affect the final error of the model estimate. We demonstrate it on the Nikon fish-eye lens model, Equation (5.33) and commonly used polynomial models 2^{nd} order $\theta = a_1 r + a_2 r^2$, and on the 3^{rd} order model $\theta = a_1 r + a_2 r^2 + a_3 r^3$. The constants a_i represent parameters of the models, r is the radius of an image point w.r.t. $(u_0, v_0)^\top$ and θ is the angle between the corresponding 3D vector and the optical axis. We used data measured on the Nikon fish-eye lens in an optical laboratory [50] as the ground truth. The uncertainty of the ground truth data measurement in the angle were $\pm 0.02°$ and in the radius were ± 0.005 mm.

We ordered ground truth points into a sequence by their radius. First, we fitted all three models to the first three (a minimal subset to compute parameters of the models) ground truth points in the sequence. We tested the fitted models on all ground truth points and computed RMS error. Then we fitted the models to the first four ground truth points and so on until all ground truth data were used for the fitting. The RMS error on all data were computed for every fit. Gaussian noise with $\sigma = 1$ pixel was added



(a) (b) (c)

Figure 5.36.: Model estimation from correspondences obtained by [58]. (a) A *wrong* model. All points were used in the model estimation by a RANSAC. The model, however, suits only to points near the center of the view field circle. Other points were marked as outliers. (b) A *correct* model. Only points near the boundary of the view field circle were used for computing the model. The model suits to points near the center as well as to points near the boundary. The circles in (a, b) encircle inliers. (c) Sampling zones. The circles show boundaries of the zones used for point sampling in the RANSAC.

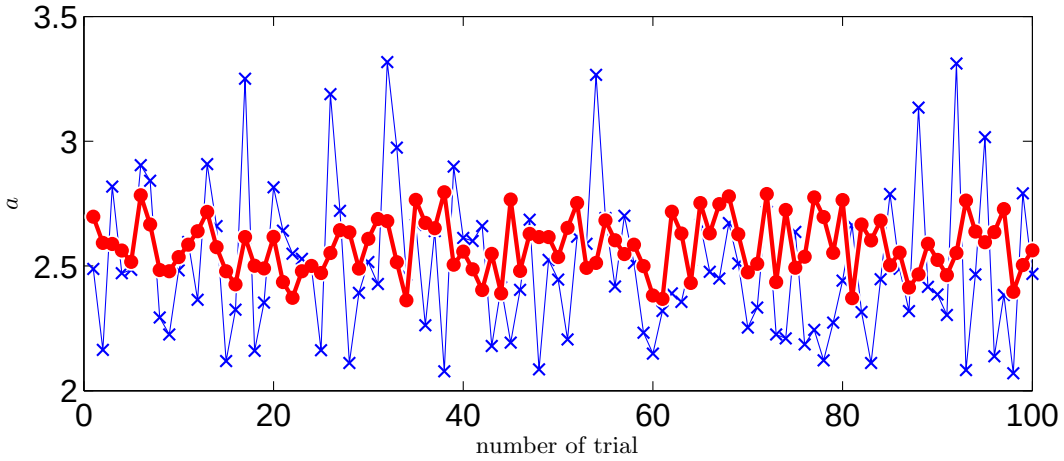


Figure 5.37.: Camera model parameter estimation without bucketing (blue thin curve with crosses) and with the bucketing technique (red thick curve with dots). The estimation procedure was run 100 times on the one image pair from Figure 5.36.

to the ground truth data and 100 trials were performed to see influence of noise on the model fitting.

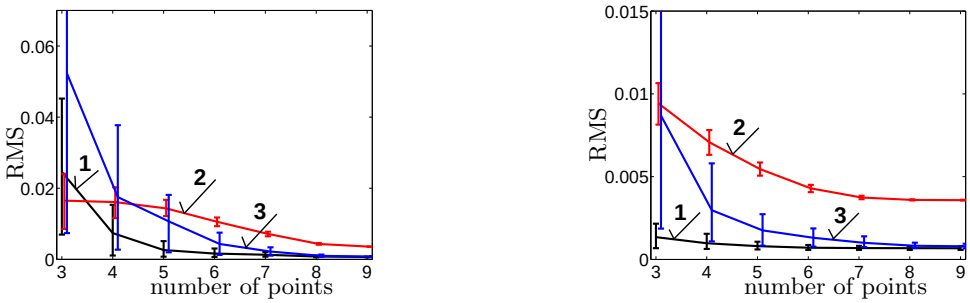
Second, we repeated the same procedure but the ground truth points were added from the end of the sequence (i.e., from the boundary to the center of the view field circle).

Figure 5.38 shows both experiments. Notice that noise affect the model fitting less when the number of points from which the model is computed increases. It can be seen from Figure 5.38(a), that RMS error is very high for the minimal set of three points and decreases significantly only when the points close to the boundary of the view field circle are included. When a minimal point sample, which is close to the view field boundary, is taken, see Figure 5.38(b), the RMS error becomes very low and adding more points that are closer to the center does not change the RMS error dramatically. It follows that the points near the boundary of the view field circle are much more important to obtain the correct model than the points near the center. It can also be seen in Figure 5.38 that the commonly used polynomial models yield larger prediction error.

By collecting the results from the two previous subsections we arrive at the conclusion: to obtain the correct model, it is necessary to exclude the points near the center $(u_0, v_0)^\top$ a priori from the RANSAC sampling and prefer the points near the boundary. Thus the points near the center are excluded from the RANSAC sampling and the rest of the image is split into three zones with equal areas from which the same number of points are randomly chosen by the RANSAC, see Figure 5.36(c). This helps to avoid the degenerate configurations, strongly biased estimates, and it decreases the number of RANSAC iterations. It can be seen in Figure 5.37 that the estimation of the camera model parameter a is more accurate (smaller deviation) by using the bucketing technique. The same result was achieved for the second camera model parameter b as well.

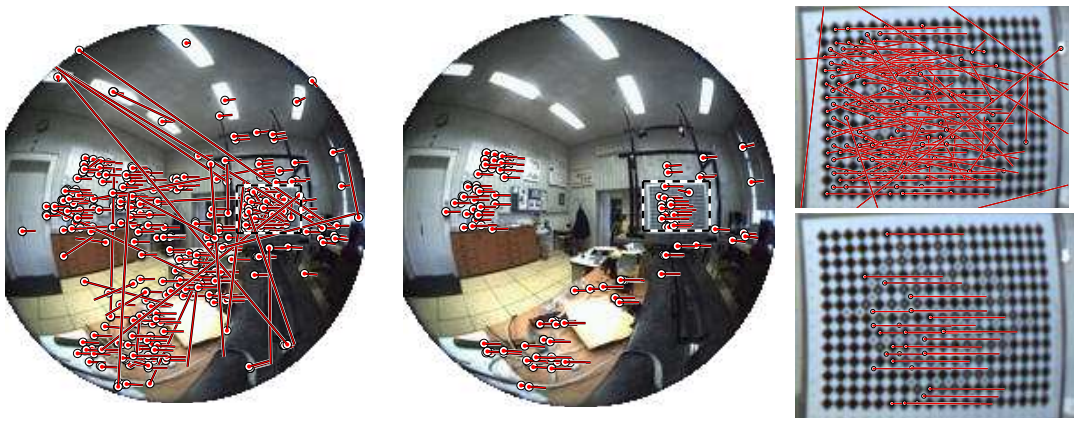
The zone sampling can be, of course, replaced by a more general method if *i)* it is assured that probability of choosing a point near the boundary is larger than the probability of choosing the point near the center $(u_0, v_0)^\top$ and *ii)* the selected points are not close to each other. The method built on this assumptions is more robust but more complex. The zone method is simple, leads to good results, and is sufficient for most of practical applications. The zone sampling (or the bucketing) technique was used in all auto-calibration experiments presented in the thesis.

The experiment in Figure 5.39 shows how the bucketing method based on the RANSAC paradigm succeeds in outlier detection. The Nikon fish-eye lens was mounted on the PixeLink CMOS camera providing 1280×1024 pxl images with diameter of the view field circle 1084 pixels. The auto-calibration method, Section 5.5.4, was applied. It is usually difficult for matching algorithms that do not use the correct omnidirectional camera model and epipolar geometry to find correct matches between images, especially in images with repetitive texture, as Figure 5.39(c) shows. Since there are many outliers, the standard RANSAC would require many iterations to provide a good estimate. As a remedy to this problem the method based on the RANSAC with bucketing can be used and unbiased camera model parameters, epipolar geometry, and inliers can be obtained in significantly fewer iterations.



(a) (b)

Figure 5.38.: Prediction error, i.e. the influence of point positions and the number of points used for model fitting. Error bars with the mean, 10th and 90th percentile values are shown. The x-axis represents the number of ground truth points used for the model fitting, the y-axis represents the RMS error on all ground truth data. (a) Points are being added from the center $(u_0, v_0)^\top$ to the boundary of the view field circle. (b) Points are being added from the boundary to the center. The proposed model (black line labelled by 1), the 2^{nd} order polynomial (red line labelled by 2), and the 3^{rd} order polynomial (blue line labelled by 3) are considered. The graphs for the 2^{nd} and the 3^{rd} order polynomials are slightly shifted to the right to make noise bars visible.



(a) (b) (c)

Figure 5.39.: Performance of the auto-calibration method based on the bucketing. (a) Tentative correspondences between a pair of the omnidirectional images found by technique [58]. Circles mark points in the first image, lines join them to their matches in the next one. (b) Detected inliers. (c) A detail view of an object with repetitive texture. Top: tentative correspondences. Bottom: inliers detected by the method.

5.7. Algorithm

All auto-calibration methods presented here are described by the following algorithm. Suppose two omnidirectional images observing the same rigid scene taken from two different positions.

1. Find an ellipse corresponding to the view field of the camera. Transform the images so that the ellipse becomes a circle, determine $\mathtt{A}$ and $\mathbf{t}$. Establish 9 (or 15 for two-parametric fish-eye (TPFE) lens model) point correspondences between two images $\{\mathbf{u}_1 \leftrightarrow \mathbf{u}_2\}$.
2. Scale the image points $\mathbf{u} := \frac{\mathbf{u}}{1000}$ to obtain better numerical stability[1].
3. Create matrices $\mathtt{D}_i$ from $\{\mathbf{u}_1 \leftrightarrow \mathbf{u}_2\}$ according to the type of a mirror or a lens and solve the PEP. Use, e.g., MATLAB: $[\mathtt{H}\ \mathbf{a}] = \mathtt{polyeig}(\mathtt{D}_1, \mathtt{D}_2, \ldots)$, $\mathtt{H}$ is the matrix with columns $\mathbf{f} \in \mathbb{R}^{9\times 1}$ (or $\mathbf{h} \in \mathbb{R}^{15\times 1}$ for TPFE), $\mathbf{a}$ is a vector with solutions of a.
4. Choose only real positive finite a's (other solutions seem never be correct), 1–4 solutions remain. For every a, there is the corresponding essential matrix $\mathtt{F}$ reshaped from vector $\mathbf{f}$ (for TPFE first nine elements of $\mathbf{h}$ give $\mathtt{F}$, last six elements give six possible b's).
5. Create 3D rays from the a (and the b) for all point correspondences. Compute the angular error, Equation (4.22), for all pairs $\{a \leftrightarrow \mathtt{F}(\leftrightarrow b)\}$ as the sum of errors for all correspondences. The pair with the minimal error is the solution, and a, (b), and the essential matrix $\mathtt{F}$ are obtained.

For integrating the algorithm into the RANSAC, 9 (or 15) points are "randomly" selected from the whole set of automatically detected correspondences and steps 1–4 are repeated until the model fitting the highest number of matches is found.

5.8. Degenerate configurations and motions

Suppose that vectors

$$\mathbf{p} = \begin{pmatrix} u \\ v \\ g(\|\mathbf{u}\|, \mathbf{a}) \end{pmatrix} \tag{5.51}$$

and

$$\bar{\mathbf{p}} = \begin{pmatrix} u \\ v \\ g(\|\mathbf{u}\|, \mathbf{a}) + d \end{pmatrix}, \qquad d \in \mathbb{R} \tag{5.52}$$

are constructed from the same pre-calibrated image point $\mathbf{u}$ and from a different third coordinate. All models presented in this chapter (PCD, HCD, SCD, fish-eye) can be rearranged to the form of Equation (5.51) where $h(\|\mathbf{u}\|) = 1$, see Equation (4.4). Assume

[1] 1000 is approximately the maximal radius of view circle. The aim is to scale the maximal radius to be ≈ 1.

that $\mathbf{p}$, $\bar{\mathbf{p}}$ are such that the epipolar geometry

$$\begin{aligned} \mathbf{p}_2^\top \, \mathtt{F} \, \mathbf{p}_1 &= 0 \,, \\ \bar{\mathbf{p}}_2^\top \, \mathtt{F} \, \bar{\mathbf{p}}_1 &= 0 \end{aligned} \tag{5.53}$$

holds for them. To simplify the situation, assume further that the $\mathtt{F}$ represents a pure translation, i.e.,

$$\mathtt{F} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & t_x \\ -t_y & -t_x & 0 \end{pmatrix} .$$

Let $w = g(\|\mathbf{u}\|, \mathbf{a})$ and re-arrange Equation (5.53) as

$$\begin{aligned} \bar{\mathbf{p}}_2^\top \, \mathtt{F} \, \bar{\mathbf{p}}_1 &= 0 \\ \left[\mathbf{p}_2^\top + \begin{pmatrix} 0 & 0 & d \end{pmatrix} \right] \mathtt{F} \left[\mathbf{p}_1 + \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix} \right] &= 0 \\ \underbrace{\mathbf{p}_2^\top \, \mathtt{F} \, \mathbf{p}_1}_{0} + \mathbf{p}_2^\top \, \mathtt{F} \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix} + \begin{pmatrix} 0 & 0 & d \end{pmatrix} \mathtt{F} \, \mathbf{p}_1 + \underbrace{\begin{pmatrix} 0 & 0 & d \end{pmatrix} \mathtt{F} \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix}}_{0} &= 0 \\ d \, \mathbf{p}_2^\top \begin{pmatrix} t_y \\ t_x \\ 0 \end{pmatrix} + d \begin{pmatrix} -t_y & -t_x & 0 \end{pmatrix} \mathbf{p}_1 &= 0 \,, \end{aligned}$$

which is for $d \neq 0$ equivalent to

$$u_2 \, t_y + v_2 \, t_x - u_1 \, t_y - v_1 \, t_x = 0 \,, \tag{5.54}$$

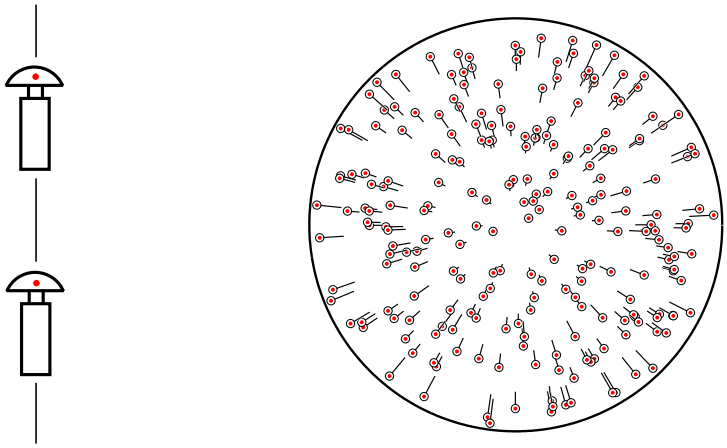
where $\mathbf{p}_1 = (u_1, \, v_1, \, w_1)^\top$ and $\mathbf{p}_2 = (u_2, \, v_2, \, w_2)^\top$.

Theorem 3 *Let the motion between two cameras be a pure translation* $\mathbf{T} = (t_x, \, t_y, \, t_z)^\top$. *If it holds for all correspondences* $(u_1, \, v_1) \leftrightarrow (u_2, \, v_2)$ *that*

$$u_2 \, t_y + v_2 \, t_x - u_1 \, t_y - v_1 \, t_x = 0 \,, \tag{5.55}$$

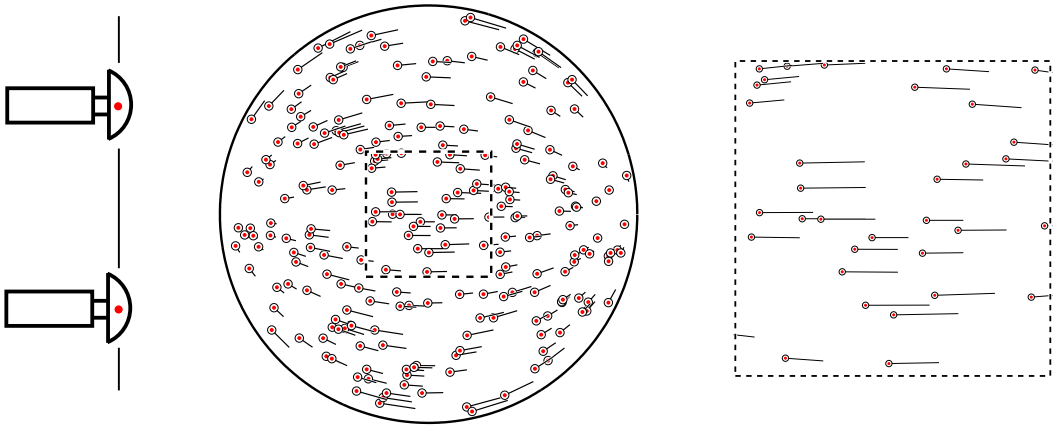
then there is an ambiguity in determination of the function $g(\|\mathbf{u}\|, \mathbf{a})$ *up to an additive element* d.

Consequence 4 *If Equation (5.55) is fulfilled, the ambiguity in the third coordinate of the vector* $\mathbf{p}$ *for pure translation causes that we estimate the vector of parameters* $\bar{\mathbf{a}}$ *instead of* $\mathbf{a}$ *in such a way that the parameters* $\bar{\mathbf{a}}$ *move the function* $g(.)$ *along the* z *axis, i.e.,* $g(\|\mathbf{u}\|, \bar{\mathbf{a}}) = g(\|\mathbf{u}\|, \mathbf{a}) + d$. *However, the essential matrix* $\mathtt{F}$ *is estimated correctly.*



(a) (b)

Figure 5.40.: Forward motion of the camera. (a) Schematic diagram. (b) Point correspondences. Circles mark points in the first image, lines join them to their matches in the next one.



(a) (b) (c)

Figure 5.41.: Side motion of the camera. (a) Schematic diagram. (b) Point correspondences. (c) Detail view.

Suppose a forward motion of the camera, i.e., the motion along the optical axis, see Figure 5.40(a). Then $t_x = t_y = 0$ and the Equation (5.55) holds. It means that there is an ambiguity in determining the parameters of the camera model and the camera cannot be calibrated from forward translation. Figure 5.40(b) shows the motion of correspondences in the image for the forward motion.

Suppose a side way motion of the camera, i.e., the motion perpendicular to the optical axis, Figure 5.41(b). Let $t_x \neq 0, t_y = 0$ then Equation (5.55) becomes

$$(v_2 - v_1)\, t_x = 0\,, \tag{5.56}$$

what is fulfilled iff $v_2 = v_1$. Figure 5.41(b) shows the motion of correspondences in the image for the side way motion. The correspondences go from one epipole to the other. In detail view, Figure 5.41(c), it can be seen that in the central part of the image the condition $v_2 = v_1$ approximately holds. It means that it is impossible to obtain correct parameters from correspondences near the center and from a side motion. However, we showed in Section 5.6 that the points near the center has no contribution to the model fitting for another reason anyway.

5.9. Closure

In this chapter, we explained the auto-calibration method for catadioptric (parabolic, hyperbolic, spherical mirror) and dioptric (Nikon, Sigma fish-eye lens) cameras. First, we showed the method for estimating radial distortion for narrow-angle view cameras. Second, we showed how the method can be generalized to omnidirectional cameras and derived models leading to the Polynomial Eigenvalue Problem for several types of the omnidirectional cameras. We designed the auto-calibration methods, based on solving the PEP, leading to fully calibrated cameras from point correspondences only. The auto-calibration methods allow simultaneous estimation of the camera model and the essential matrix together with validated point correspondences. We showed the performance on real data for each omnidirectional camera. We proposed the method for speeding up the convergence of a RANSAC-based estimation technique. Finally, we investigated some situations where the calibration can be performed only up to an ambiguity.

As a result, the parabolic mirror has an important advantage over the hyperbolic, the spherical mirror and fish-eye lenses. The parabolic mirror directly leads to the PEP, the hyperbolic, spherical mirrors and fish-eye lenses do not. To obtain the PEP for other central omnidirectional cameras a suitable linearization of the camera model has to be done.

6 Real Non-Central Catadioptric Cameras

In the previous chapter, we assumed that catadioptric cameras consisting of a parabolic and a hyperbolic mirror can be aligned in such a way that they possess a single projection center. We also approximated the spherical catadioptric camera to get a fictive projection center. Such catadioptric cameras can be treated as central cameras and the theory of central catadioptric cameras [3, 10, 36, 39, 85] can be employed.

In practice, however, catadioptric cameras are not central. The most common reasons are that *i)* a non-telecentric lens is used for a parabolic mirror or the lens and the parabolic mirror axes are not aligned properly, *ii)* a perspective camera is not placed in one of the focal points of a hyperbolic or an elliptical mirror, or *iii)* the mirror shapes, e.g., the spherical or uniform resolution mirrors [46, 32], are designed so that they do not posses a single viewpoint property. All the above may cause that the catadioptric camera becomes non-central and there is no single viewpoint from which all rays would emanate.

Using the central camera model for a non-central camera leads to an inaccurate determination of 3D rays corresponding to an image point and consequently to a skewed 3D reconstruction as Figure 6.1(top left) shows. A remedy is to derive the correct non-central camera model capturing the true distribution of reflected rays. See Figure 6.1 to compare how the 3D reconstruction was improved by using the non-central model.

To build a 3D reconstruction from automatically established point correspondences contaminated by mismatches, we need to be able to incorporate the auto-calibration procedure into some robust estimation technique, e.g., RANSAC. However, there is a price to pay for the accuracy. A non-central model is often very complex with a large number of parameters and it is often impossible (or computationally very complex) to use the non-central model in a RANSAC-based estimation of multiple-view geometry. Thus, we are facing a problem of automatic & accurate reconstruction: "To be accurate, we have to use a complex non-central model. To be automatic, we cannot use a complex model."

We show in this chapter how to solve the above problem by constructing a hierarchy of camera models. The hierarchy starts with a simplified model that is accurate enough to distinguish between correct (inliers) and incorrect (outliers) correspondences and simple enough to provide a tractable optimization problem when using the RANSAC. In all cases presented here, the simplified models, derived in Chapter 5, are central. The simplified models allow to recover inliers automatically. Secondly, an accurate and complex non-central model is fitted to the correspondences validated in the first step. The non-central

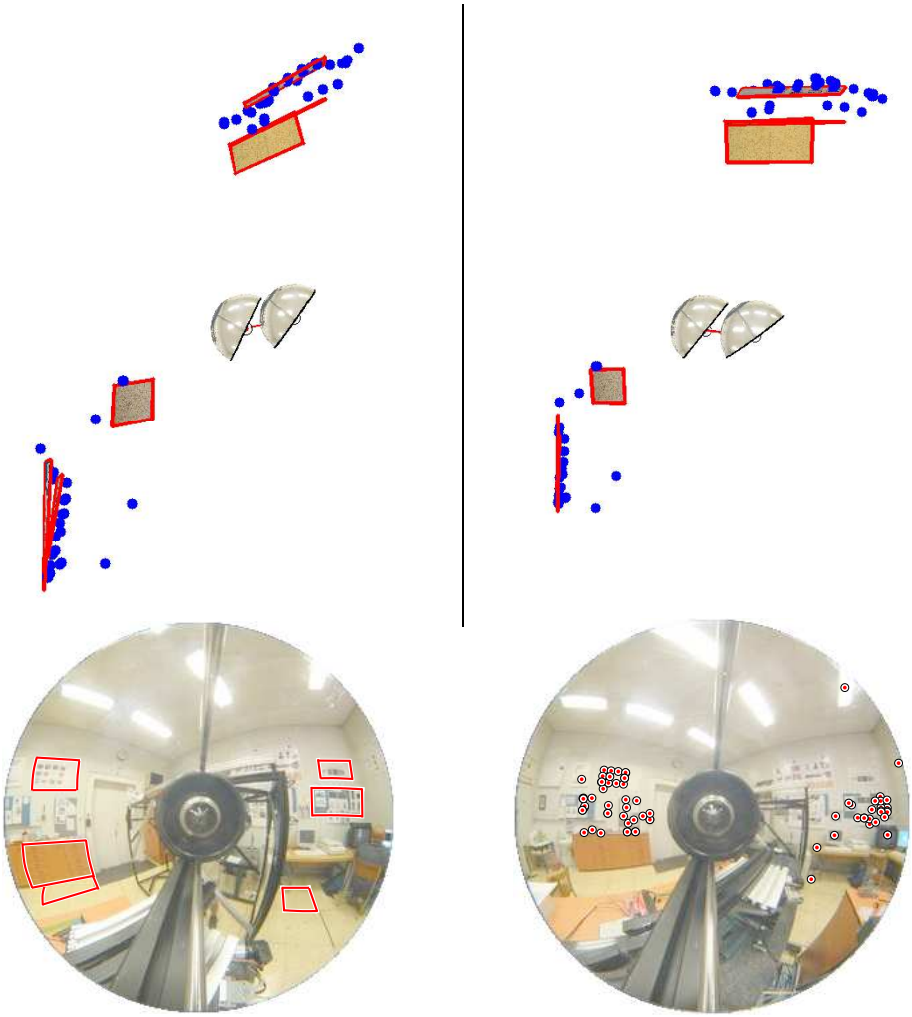


Figure 6.1.: The top views of the 3D reconstructions of a rectangular room. Top Left: A skewed reconstruction. A central model for a real para-catadioptric (PCD) camera was used. Top Right: The correct reconstruction. A non-central PCD camera model was used. Bottom: Input image pair showing manually marked rectangular regions and point matches established automatically by [58].

model allows to reconstruct scene geometry accurately. In general, the art is to find *i)* a simplified model that is accurate enough to reject outliers but simple enough to provide a tractable RANSAC estimation problem and *ii)* a sufficiently accurate non-central model providing accurate 3D reconstruction from correct image correspondences.

We present a technique allowing to solve the correspondence problem, auto-calibrate cameras, and compute a 3D metric reconstruction automatically from two uncalibrated non-central catadioptric images. We use the result (auto-calibration methods) from Chapter 5. The technique is demonstrated on most common (parabolic, hyperbolic, and spherical) quadric mirrors.

6.1. Non-central PCD camera

An ideal PCD camera consisting of a properly assembled a parabolic mirror and a telecentric camera possesses the central projection, hence it has a single viewpoint. In practice, however, a standard perspective camera with a large focal length lens is used instead of an ideal telecentric lens and moreover, the optical axis of the perspective camera is not aligned with the mirror axis of symmetry exactly. Both, non-telecentric lens and inaccurate alignment of the perspective camera with the mirror, cause that the PCD camera becomes non-central and does not possess a single viewpoint.

6.1.1. Non-central model

Let us derive a general imaging model for a non-central para-catadioptric camera. The technique of derivation is in some steps similar to the derivation of the central catadioptric camera models [85]. General quadric surface can be expressed as

$$(\mathbf{x}^\top \quad 1)\,\mathtt{Q}\begin{pmatrix}\mathbf{x}\\1\end{pmatrix} = 0\,, \tag{6.1}$$

where $\mathtt{Q}$ is a 4×4 matrix that varies depending on the type of the quadric, $\mathbf{x}$ from $\mathbb{R}^3$ is a point on the quadric surface. The matrix $\mathtt{Q}$ for a paraboloid of revolution, Equation (5.7), in the coordinate system placed to the focus of the paraboloid $\mathbf{F}$ is

$$\mathtt{Q} = \begin{bmatrix} -\frac{1}{a^2} & 0 & 0 & 0\\ 0 & -\frac{1}{a^2} & 0 & 0\\ 0 & 0 & 0 & -\frac{1}{a}\\ 0 & 0 & -\frac{1}{a} & 1 \end{bmatrix}, \tag{6.2}$$

where, remind, a is twice the distance from the vertex to the focal point and is called the mirror parameter.

Suppose a Cartesian world coordinate system W. The Cartesian coordinate system of a mirror, placed in $\mathbf{F}$, is rotated and translated by $\mathtt{R}_m \in \mathbb{R}^{3\times3}$ and $\mathbf{t}_m \in \mathbb{R}^{3\times1}$ w.r.t. W. The Cartesian coordinate system of a perspective camera, placed in the optical center $\mathbf{C}$, is related to the mirror coordinate system by $\mathtt{R}_c$ and $\mathbf{t}_c$, see Figure 6.2.

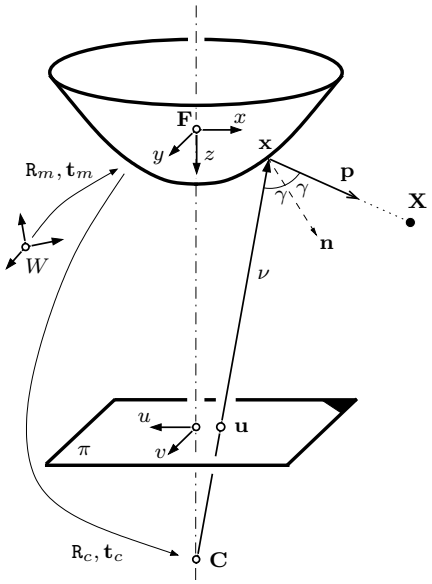


Figure 6.2.: Coordinate system of the para-catadioptric camera.

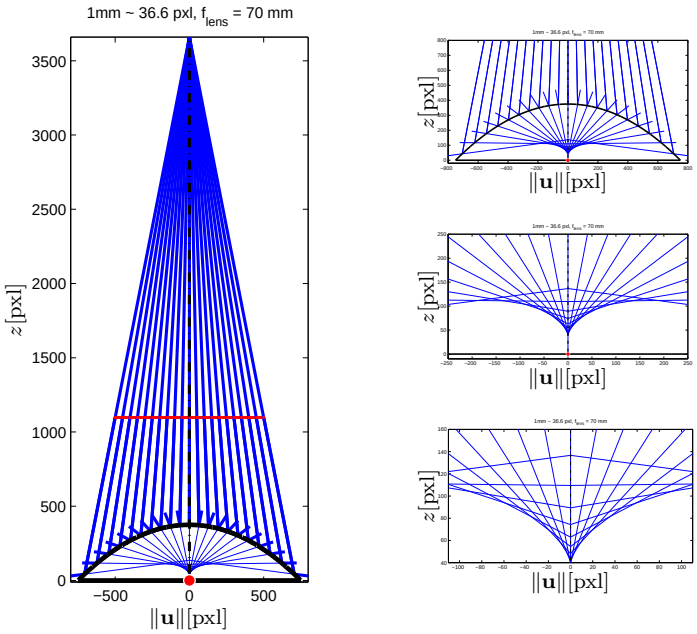


Figure 6.3.: The caustic. Rays reflected by the mirror are tangent to a curve, called the caustic. Left: Standard camera looking at the parabolic mirror. Right: Detail views of the caustic.

From image point to its projection ray

A conventional perspective camera projects a 3D point $\mathbf{x}$ on a mirror surface into the image plane by the standard projection equation [43]

$$\mathbf{u} = \frac{1}{\lambda}\mathtt{K}\,\mathtt{R}_c[\mathtt{I}|-\mathbf{t}_c]\begin{pmatrix}\mathbf{x}\\1\end{pmatrix}, \tag{6.3}$$

where $\mathtt{K}$ is a 3×3 upper triangular camera calibration matrix [43], $\mathbf{u}\in\mathbb{R}^3$ and $\lambda\in\mathbb{R}\backslash\{0\}$.

The line ν emanating from the camera center $\mathbf{C}$ in the direction $\mathbf{u}$ can be parametrized by λ and written, using Equation (6.3), as

$$\begin{aligned}\nu &= \lambda\begin{pmatrix}r\\s\\t\end{pmatrix}+\begin{pmatrix}k\\l\\m\end{pmatrix}=\lambda\mathbf{v}+\mathbf{t}_c=\\ &= \lambda\mathtt{R}_c^\top\mathtt{K}^{-1}\mathbf{u}+\mathbf{t}_c\,.\end{aligned} \tag{6.4}$$

The intersection of the line ν with the mirror gives a mirror point $\mathbf{x}$. The intersection is obtained by substituting Equation (6.4) and Equation (6.2) into the mirror Equation (6.1) what yields

$$(-r^2-s^2)\lambda^2+2(-kr-at-ls)\lambda+(a^2-k^2-l^2-2am)=0\,. \tag{6.5}$$

Quadratic equation (6.5) gives two solutions for λ, i.e.,

$$\begin{aligned}\lambda_{1,2}=\frac{1}{(-r^2-s^2)}\Big(&kr+at+ls\pm\\ &\pm\sqrt{(kr+at+ls)^2+(r^2+s^2)(a^2-k^2-l^2-2am)}\Big)\,.\end{aligned} \tag{6.6}$$

Going from $\mathbf{C}$ in the direction of $\mathbf{u}$, we are interested in the intersection which is closer to the point $\mathbf{C}$, hence the smaller λ is chosen.

Substituting λ from Equation (6.6) into Equation (6.4) yields the point on the mirror surface. This point $\mathbf{x}=\lambda\mathtt{R}_c^\top\mathtt{K}^{-1}\mathbf{u}+\mathbf{t}_c$ is expressed in the world coordinate system W as

$$\mathbf{x}_w=\mathtt{R}_m^{-1}\mathbf{x}+\mathbf{t}_m=\mathtt{R}_m^{-1}\left(\lambda\mathtt{R}_c^\top\mathtt{K}^{-1}\mathbf{u}+\mathbf{t}_c\right)+\mathbf{t}_m\,. \tag{6.7}$$

The ray ν is reflected at the point $\mathbf{x}$ by the mirror such that the angles γ of incident and coincident rays to the surface normal are equal. The normal to the mirror surface in the mirror point $\mathbf{x}$, i.e., $\mathbf{n}=(\frac{\partial z}{\partial x},\frac{\partial z}{\partial y},-1)^\top$, is given by

$$\mathbf{n}=\left(\frac{-x}{a},\frac{-y}{a},-1\right)^\top. \tag{6.8}$$

If the normal vector $\mathbf{n}$ is normalized to $\hat{\mathbf{n}} = \frac{\mathbf{n}}{\|\mathbf{n}\|}$ of unit length, the directional vector of the reflected ray $\mathbf{p}$ can be computed as

$$\begin{aligned} \mathbf{p} &= \mathbf{v} - 2(\mathbf{v} \cdot \hat{\mathbf{n}}) \cdot \hat{\mathbf{n}} = \\ &= \mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u} - 2\left(\left(\mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u}\right) \cdot \hat{\mathbf{n}}\right) \cdot \hat{\mathbf{n}}\,. \end{aligned} \tag{6.9}$$

The vector $\mathbf{p}$, expressed in the world coordinate system, i.e. $\mathbf{p}_w = \mathtt{R}_m^{-1}\mathbf{p}$, is given by

$$\mathbf{p}_w = \mathtt{R}_m^{-1}\left(\mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u} - 2\left(\left(\mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u}\right) \cdot \frac{\mathbf{n}}{\|\mathbf{n}\|}\right) \cdot \frac{\mathbf{n}}{\|\mathbf{n}\|}\right). \tag{6.10}$$

Equation (6.7) and Equation (6.10) represent the complete mapping from an image point $\mathbf{u}$ in the camera coordinate system to a pair $(\mathbf{x}_w; \mathbf{p}_w)$ in the world coordinate system. The pair $(\mathbf{x}_w; \mathbf{p}_w)$ consists of the point $\mathbf{x}_w$ on the mirror surface and the directional vector $\mathbf{p}_w$ pointing towards a scene point $\mathbf{X}$. A simulation of the derived model shows the corresponding caustic, see Figure 6.3.

Projection of a scene point to the image

Given a scene point $\mathbf{X}$, there is no direct solution for its image projection $\mathbf{u}$. It can be easily designed as an iterative method incorporating the mapping from an image point to its projection ray (derived in the previous subsection).

First, the initial image $\mathbf{u}$ of the $\mathbf{X}$ is obtained using a central model (derived later), for which a direct solution exists. Second, the iterative method is used. The iterative method minimizes (over coordinates of $\mathbf{u}$) the distance between a ray, computed from $\mathbf{u}$ by the non-central model, and the 3D point $\mathbf{X}$. The method converges very quickly thanks to a good initial estimate of $\mathbf{u}$ provided by the approximate central model.

6.1.2. Central model

A non-central model can be approximated by a central one in order to find a sufficiently simple model which is accurate enough to distinguish between correct (inliers) and incorrect (outliers) point matches and which can be used to obtain an initial estimate of camera motion.

From image point to its projection ray

The derivation of the ideal PCD camera projection from image points to projection rays is straightforward since the central model is a special case of the non-central one. Suppose that the optical axis of the perspective camera goes through the center of the mirror, i.e., $\mathtt{R}_c = \mathrm{diag}(-1, 1, -1)$ (image is mirrored) and $\mathbf{t}_c = (0, 0, \infty)^\top$ (the camera optical center lies at infinity). Let move the world coordinate system to the coordinate system of the mirror (with the center $\mathbf{F}$), i.e., $\mathtt{R}_m = \mathrm{diag}(1, 1, 1)$ and $\mathbf{t}_m = (0, 0, 0)^\top$. Then, the vector $\mathbf{p}_w$ from Equation (6.10) becomes the same as the one which we used in Equation (5.8).

Projection of a scene point to the image

Let $\mathbf{X} \in \mathbb{R}^3$ be a scene point, and $\mathtt{R}_m$, $\mathbf{t}_m$, $\mathtt{R}_c$, $\mathbf{t}_c$, $\mathtt{K}$ be as before. The intersection of the ray going from $\mathbf{X}$ towards a center $\mathbf{F}$ with the mirror surface, Equation (6.2), can be computed from Equation (6.1)

$$\begin{bmatrix} \lambda[\mathtt{R}_m, -\mathtt{R}_m\mathbf{t}_m]\mathbf{X} \\ 1 \end{bmatrix}^\top \mathtt{Q} \begin{bmatrix} \lambda[\mathtt{R}_m, -\mathtt{R}_m\mathbf{t}_m]\mathbf{X} \\ 1 \end{bmatrix} = 0$$

as the smaller λ from the two solutions of

$$(-s^2 - r^2)\lambda^2 - 2ta\lambda + a^2 = 0\,, \tag{6.11}$$

where $(r, s, t)^\top = [\mathtt{R}_m, -\mathtt{R}_m\mathbf{t}_m]\mathbf{X}$. The point on the mirror surface, written in the camera coordinate system,

$$\mathbf{x}_c = (x_c, y_c, z_c)^\top = \mathtt{R}_c\left(\lambda[\mathtt{R}_m, -\mathtt{R}_m\mathbf{t}_m]\mathbf{X} - \mathbf{t}_c\right) \tag{6.12}$$

is projected to the image as

$$\mathbf{u} = \mathtt{K}\frac{1}{z_c}\mathbf{x}_c\,. \tag{6.13}$$

Equation (6.13) allows to compute the perspective image $\mathbf{u}$ of the scene points $\mathbf{X}$.

6.1.3. 3D reconstruction

Every i^{th} point correspondence can be represented by $(\mathbf{x}^i_{w1}; \mathbf{p}^i_{w1})$ in the first camera and by $(\mathbf{x}^i_{w2}; \mathbf{p}^i_{w2})$ in the second one, computed from Equations (6.7) and (6.10). The reconstructed 3D point $\mathbf{X}$ for one point match $(\mathbf{x}^i_{w1}; \mathbf{p}^i_{w1}) \leftrightarrow (\mathbf{x}^i_{w2}; \mathbf{p}^i_{w2})$ is obtained as the point in the center of the shortest transversal of the respective projection rays

$$d_i = \frac{|\left(\mathbf{x}^i_{w1} - \mathbf{x}^i_{w2}\right) \cdot \left(\mathbf{p}^i_{w1} \times \mathbf{p}^i_{w2}\right)|}{|\mathbf{p}^i_{w1} \times \mathbf{p}^i_{w2}|}\,. \tag{6.14}$$

The final *metric* reconstruction $\mathcal{R}_M$ is obtained by minimizing the sum of squared lengths of the shortest transversals

$$\mathcal{R}_M = \underset{\mathbf{a},\mathtt{R}_c,\mathbf{t}_c,\mathtt{R}_m,\mathbf{t}_m,\mathtt{K}}{\operatorname{argmin}} \sum_{i=1}^{N} d_i^2\,, \tag{6.15}$$

where N is the number of point matches. This is almost equivalent to minimizing the image reprojection errors since our cameras are calibrated.

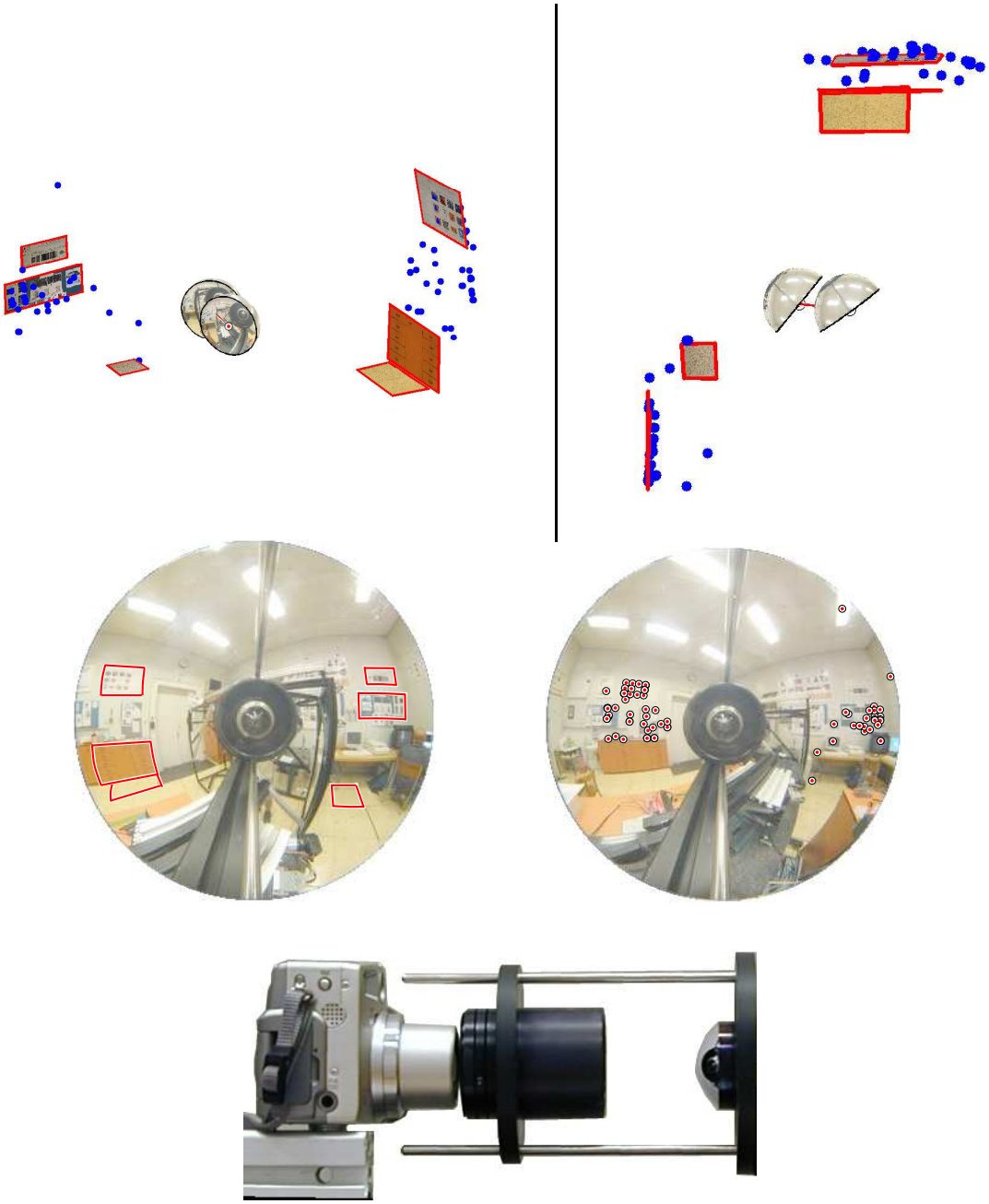


Figure 6.4.: Top left: 3D metric reconstruction from two uncalibrated para-catadioptric images. T op right: Bird's eye view of the room. Notice that the reconstructed plan is really rectangular. Middle row: The input image pair together with manually marked rectangular regions and automatically detected point correspondences [58]. Bottom: A real non-central PCD camera.

6.1.4. Experiment

We acquired two images by a real PCD camera (the same parabolic mirror as in experiment in Section 5.2.3 and Canon PowerShot G2 digital camera with the diameter of the view field circle 1474 pixels $\sim 210°$).

The tentative correspondences were obtained by [58]. As a result of applying the calibration method described in Section 5.2.2, an initial camera model and the essential matrix were obtained and outliers were rejected.

The camera model parameter and relative camera position obtained from the essential matrix were used to perform an initial 3D reconstruction as Section 4.5 describes. The 3D reconstruction was improved by the non-linear bundle adjustment, Equation (6.15), tuning the second camera matrix $\mathtt{P}_2 = [\mathtt{R}_m \,|\, \mathtt{t}_m]$ (the first camera was fixed to $\mathtt{P}_1 = [\mathtt{I} \,|\, \mathbf{0}]$) and the internal camera model parameters, same for both cameras. The internal camera model parameters are *i)* the center of the radial symmetry in the image $(u_0, v_0)^\top$, i.e., the intersection of the mirror axis with the image plane, *ii)* the mirror shape parameter a, *iii)* the perspective camera orientation and position w.r.t. the mirror coordinate system $\mathtt{R}_c$, $\mathbf{t}_c$, and *iv)* the focal length f of the perspective camera. The square pixel and zero skew of the perspective camera were assumed.

To show the quality of the 3D reconstruction, some correspondences, like corners on the walls, have been established manually. The estimated camera matrices were used to reconstruct these points. See Figure 6.4 for the result.

6.2. Non-central HCD camera

An ideal hyperbolic catadioptric camera properly assembled from a hyperbolic mirror and a perspective camera placed in the second focal point of the hyperboloid realizes the central projection, hence it possesses a single viewpoint. In practice, however, the optical center of a perspective camera is often not placed exactly in the second focal point of the hyperboloid. Inaccurate alignment of the perspective camera with the mirror causes that the catadioptric camera becomes non-central and does not possess a single viewpoint.

6.2.1. Non-central model

The derivation of a non-central hyperbolic catadioptric model is analogical to derivation of a parabolic mirror model, therefore only final equations are presented.

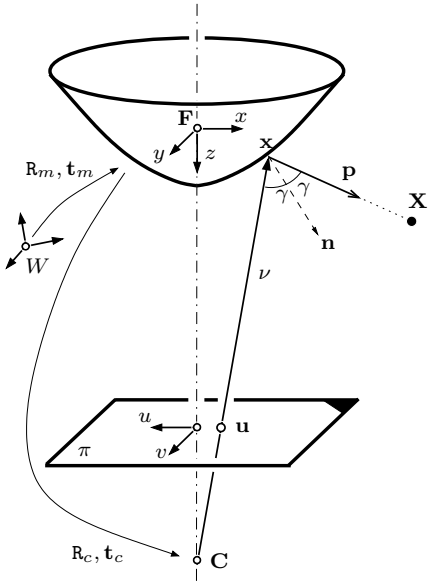


Figure 6.5.: Coordinate system of the hyperbolic catadioptric camera.

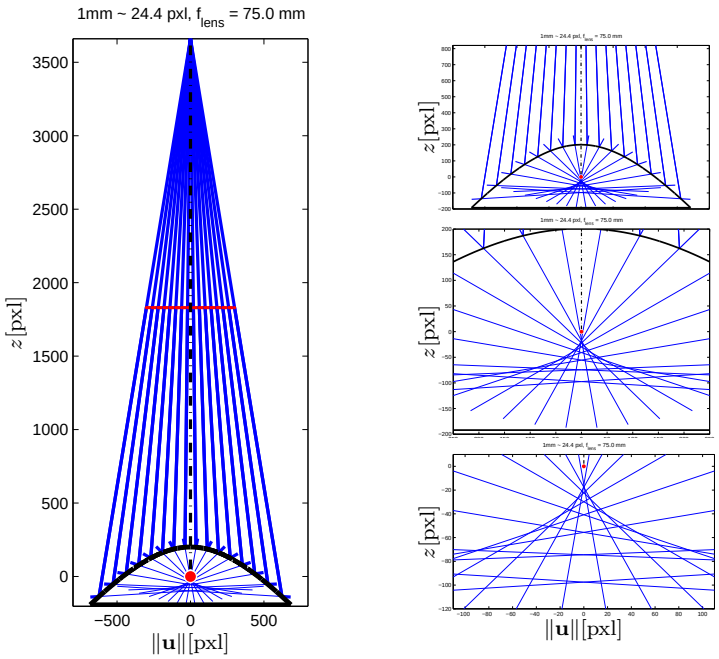


Figure 6.6.: The caustic. Left: Perspective camera looking at the hyperbolic mirror. Right: A detail views of the caustic.

From image point to its projection ray

The matrix $\mathtt{Q}$ for a hyperboloid of revolution with Equation (5.13) in the coordinate system according to Figure 6.5 with the center in $\mathbf{F}$ is

$$\mathtt{Q} = \begin{bmatrix} -\frac{1}{b^2} & 0 & 0 & 0 \\ 0 & -\frac{1}{b^2} & 0 & 0 \\ 0 & 0 & \frac{1}{a^2} & -\frac{e}{a^2} \\ 0 & 0 & -\frac{e}{a^2} & \frac{b^2}{a^2} \end{bmatrix}, \tag{6.16}$$

where a, b, $e = \sqrt{a^2 + b^2}$ are the parameters of a hyperboloid.

Substituting Equation (6.4) and Equation (6.16) into the mirror equation (6.1) yields

$$\begin{aligned}(t^2b^2 - s^2a^2 - r^2a^2)\lambda^2 + 2(mtb^2 - etb^2 - lsa^2 - 2kra^2)\lambda + \\ + (m^2b^2 - k^2a^2 - 2meb^2 - l^2a^2 + b^4) = 0\,. \end{aligned} \tag{6.17}$$

Quadratic equation (6.17) gives two solutions for λ. The smaller one represents the correct intersection of the ray with the mirror surface. Substituting λ from Equation (6.17) into Equation (6.4) yields the point on the mirror surface, i.e., the point $\mathbf{x} = \lambda \mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u} + \mathbf{t}_c$. The point $\mathbf{x}$, expressed in the world coordinate system, is given by Equation (6.7).

The normal vector to the hyperbolic mirror surface in a point (x, y) is given by

$$\mathbf{n} = \left(-\frac{a}{b}\frac{x}{\sqrt{b^2 + x^2 + y^2}}, -\frac{a}{b}\frac{y}{\sqrt{b^2 + x^2 + y^2}}, -1 \right)^\top. \tag{6.18}$$

The vector $\mathbf{p}_w$ in the world coordinate system W emanating from the mirror surface point $\mathbf{x}_w$ to a scene point is given by Equation (6.10), in which the normal $\mathbf{n}$ is substituted from Equation (6.18).

A simulation of the non-central hyperbolic catadioptric camera, using the derived model, shows the shape of the corresponding caustic, see Figure 6.6.

Projection of a scene point to the image

The projection of a scene point to the image can be obtained by an iterative method, analogical to the one used for the PCD camera. The initial estimate is obtained by using the central model of the HCD camera which follows.

6.2.2. Central model

From image point to its projection ray

The derivation of the central HCD camera projection from image points to projection rays is straightforward again. Suppose that the optical axis of the perspective camera

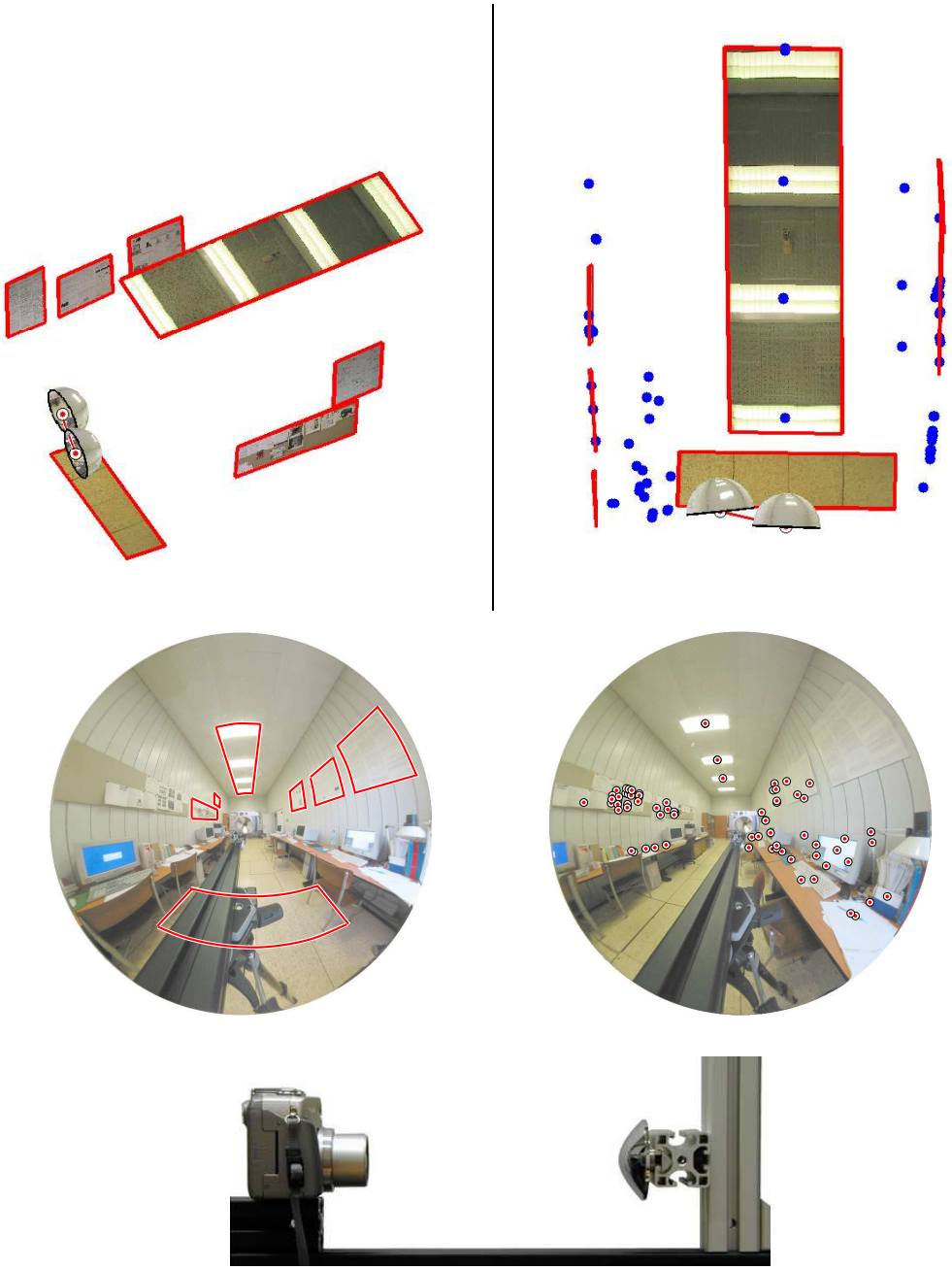


Figure 6.7.: Top left: 3D metric reconstruction from two uncalibrated hyperbolic catadioptric images. Top right: Top view of the room. Notice that the reconstructed plan is really rectangular. Middle row: The Input image pair together with manually marked rectangular regions and automatically detected point correspondences [58]. Bottom: A real non-central hyperbolic catadioptric camera.

goes through the center of the mirror, i.e. $\mathtt{R}_c = \mathrm{diag}(-1, 1, -1)$ (image is mirrored) and $\mathbf{t}_c = (0, 0, 2e)^\top$ (camera optical center is in the second focal point of the hyperboloid). Let move the world coordinate system to the coordinate system of the mirror (with the center $\mathbf{F}$), i.e., $\mathtt{R}_m = \mathrm{diag}(1, 1, 1)$ and $\mathbf{t}_m = (0, 0, 0)^\top$. Then, the vector $\mathbf{p}_w$ is the same as in Equation (5.19).

Projection of a scene point to the image

The derivation of the central HCD camera projection of a scene point to the image is analogical to the PCD camera, only λ in Equation (6.12) is computed from

$$(b^2t^2 - a^2r^2 - a^2s^2)\lambda^2 + 2e\,t\,b^2\lambda + b^4 = 0\,. \tag{6.19}$$

The decision which λ is the solution is the following: if $\lambda_{1,2}$ have opposite signs then $\lambda = \max(\lambda_{1,2})$ otherwise $\lambda = \min(\lambda_{1,2})$.

6.2.3. Experiment

We acquired two images by a hyperbolic catadioptric camera. The same hyperbolic mirror as in experiment in Section 5.3.3 and Canon PowerShot G2 digital camera with the diameter of the view field circle 1560 pixels $\sim 186^\circ$ were used.

The technique of the 3D reconstruction is analogical to one used for the PCD camera described in Section 6.1.4. The difference is just in the first step - outliers detection and motion estimation. It is done by the technique described in Section 5.3.2, which is designed for central hyperbolic catadioptric cameras. The auto-calibration provides validated point correspondences, a good initial estimate of the parameter c, and relative camera positions. The non-linear bundle adjustment, same as for the parabolic mirror, using a non-central HCD camera model from Section 6.2.1 is applied to build a 3D metric reconstruction. See Figure 6.7 for the final 3D metric reconstruction of the surrounding scene together with camera positions.

6.3. Non-central SCD camera

The spherical mirror is a surface that does not possess a single projection center when observed by any central (perspective or orthographic) camera. Therefore, to build accurate 3D reconstruction from spherical catadioptric images, the non-central model has to be used.

6.3.1. Non-central model

The derivation of the non-central spherical catadioptric model is analogical to the parabolic and the hyperbolic mirror, therefore it will be presented very briefly.

From image point to its projection ray

The matrix $\mathtt{Q}$ for a sphere with the equation

$$x^2 + y^2 + z^2 = a^2$$

in the coordinate system according to Figure 6.8 with the center in $\mathbf{F}$ is

$$\mathtt{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -a^2 \end{bmatrix}, \tag{6.20}$$

where a is the radius of the sphere.

Substituting Equation (6.4) and Equation (6.20) into the mirror equation (6.1) yields the quadratic equation

$$(r^2 + s^2 + t^2)\lambda^2 + 2(ls + kr + mt)\lambda + (k^2 - a^2 + m^2 + l^2) = 0\,, \tag{6.21}$$

which gives two solution for λ. The smaller one represents the correct intersection of the ray with the mirror surface. Substituting λ from Equation (6.21) into Equation (6.4) yields the point on the mirror surface, i.e. the point $\mathbf{x} = \lambda \mathtt{R}_c^\top \mathtt{K}^{-1}\mathbf{u} + \mathbf{t}_c$. The point $\mathbf{x}$ expressed in the world coordinate system is given by Equation (6.7).

The normal vector to the spherical mirror surface in a point (x, y) is given by

$$\mathbf{n} = \left(\frac{-x}{\sqrt{a^2 - x^2 - y^2}}, \frac{-v}{\sqrt{a^2 - x^2 - y^2}}, -1 \right)^\top .$$

The vector $\mathbf{p}_w$ in the world coordinate system W emanating from the mirror surface point $\mathbf{x}_w$ to a scene point is given by Equation (6.10), in which the normal $\mathbf{n}$ is substituted from Equation (6.3.1).

A simulation of the non-central spherical catadioptric camera, using the derived model, shows the shape of the corresponding caustic, see Figure 6.9.

Projection of a scene point to the image

The projection of a scene point to the image can be obtained by an iterative method, analogical to the one used for the PCD camera. The initial estimate is obtained by using the central model of the SCD camera which follows.

6.3.2. Central model

From image point to its projection ray

The derivation of the approximate central SCD camera projection from image points to projection rays is not so straightforward as it was for the PCD and the HCD cameras.

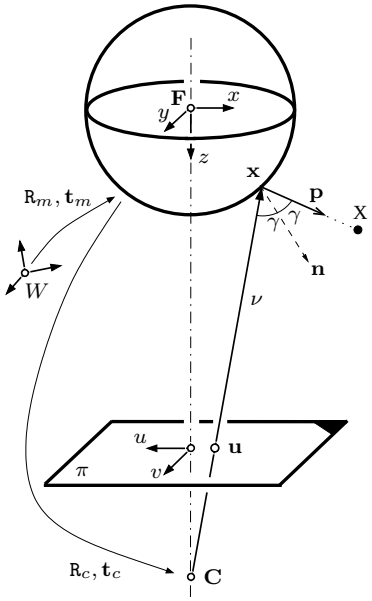


Figure 6.8.: Coordinate system of the spherical catadioptric camera.

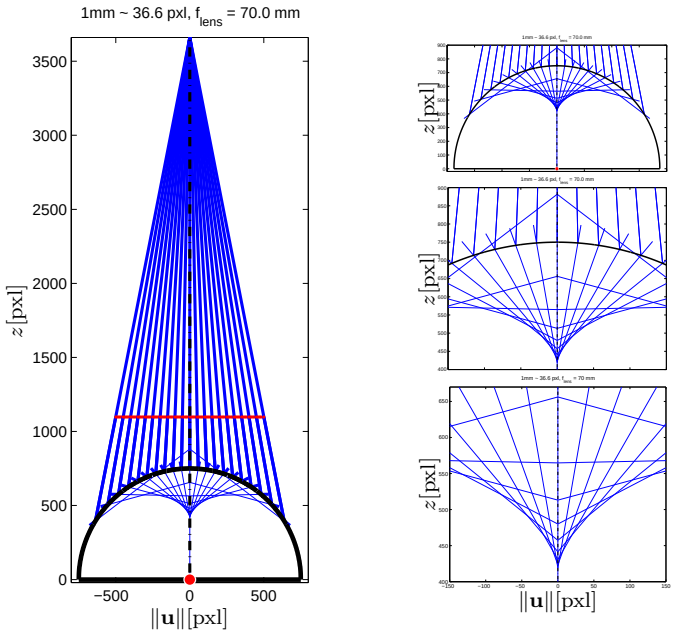


Figure 6.9.: The caustic. Left: Perspective camera looking at the spherical mirror. Right: A detail views of the caustic.

The approximate central model will have the projection center in a fictive center $\mathbf{F}_f$, which differs from the $\mathbf{F}$, see Figure 5.19.

Suppose that the optical axis of the perspective camera passes through the center of the mirror, i.e., $\mathtt{R}_c = \mathrm{diag}(-1, 1, -1)$ (image is mirrored) and $\mathbf{t}_c = (0, 0, m)^\top$. Let us move the world coordinate system to the coordinate system of the mirror (with the fictive center $\mathbf{F}_f$), i.e., $\mathtt{R}_m = \mathrm{diag}(1, 1, 1)$ and $\mathbf{t}_m = (0, 0, 0)^\top$. The matrix $\mathtt{Q}$ of a sphere given by the Equation (5.25) with the origin in the fictive point $\mathbf{F}_f$ becomes

$$\mathtt{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & k\,a \\ 0 & 0 & k\,a & (k\,a)^2 - a^2 \end{bmatrix}, \tag{6.22}$$

where $k\,a$ is the distance of $\mathbf{F}_f$ w.r.t. $\mathbf{F}$, see Figure 5.19. Substituting Equation (6.4) and Equation (6.22) into the mirror Equation (6.1) yields the vector $\mathbf{p}$, used in Equation (5.28), going towards its scene point.

Projection of a scene point to the image

The derivation of the approximate central SCD camera projection of a scene point to the image is analogical to the PCD camera, only λ in Equation (6.12) is computed from

$$(r^2 + s^2 + t^2)\lambda^2 + 2\,t\,k\,a\lambda + k^2a^2 - a^2 = 0\,. \tag{6.23}$$

6.3.3. Experiment

We acquired two images by a spherical catadioptric camera (a gear ball and Canon PowerShot G2 digital camera with the diameter of the view field circle equal to 1042 pixels $\sim 252^\circ$).

The technique of the 3D reconstruction is analogical to the PCD camera described in Section 6.1.4. The difference is just in the first step - outliers detection and motion estimation. It is done by the technique described in Section 5.4.2, which is designed for approximately central spherical catadioptric cameras. The non-linear bundle adjustment, the same as for the parabolic mirror, using the non-central SCD camera model from Section 6.3.1 is applied to build the 3D metric reconstruction. See Figure 6.10 for the final 3D metric reconstruction of the surrounding scene together with camera positions.

Results in Figure 6.4, Figure 6.7, and Figure 6.10 show that the method leads to accurate 3D metric reconstructions what can easily be verified on the top views. The reconstructed rooms are indeed rectangular. Notice, e.g., for a spherical mirror, 1024 pxl for the FOV 252° corresponds to 182 pxl for the FOV 45° of a typical perspective camera. Even though the omnidirectional cameras has approximately 5 times lower resolution a very accurate and complete 3D reconstruction can be built.

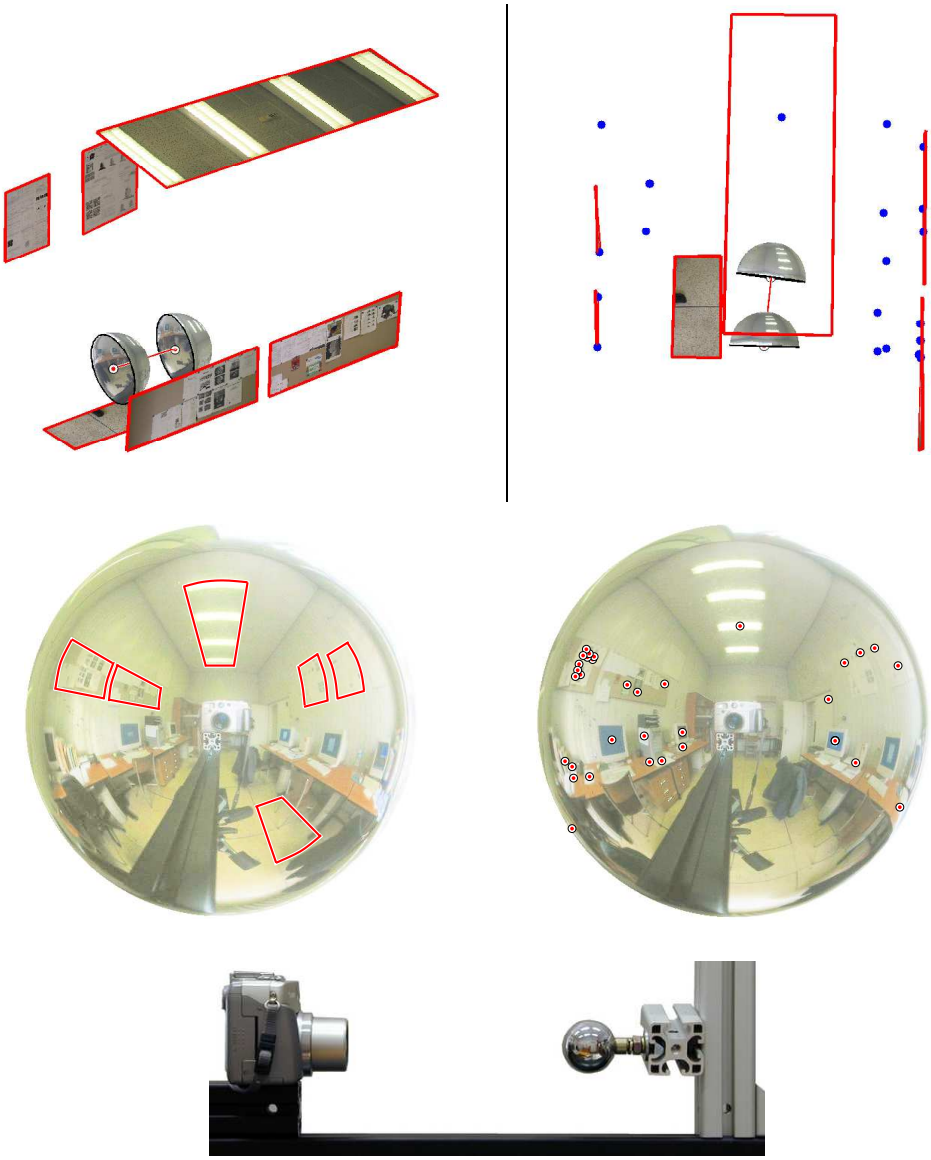


Figure 6.10.: Top left: 3D metric reconstruction from two uncalibrated spherical catadioptric images. Top right: Top view of the room. Notice that the reconstructed plan is really rectangular. Middle row: The input image pair together with manually marked rectangular regions and automatically detected point correspondences [58]. Bottom: A real non-central spherical catadioptric camera.

6.4. Algorithm

The method for the 3D reconstruction from real (non-central) catadioptric images can be described by the following algorithm. Suppose two non-central catadioptric images observing the same rigid scene taken from two different positions.

1. Apply a particular auto-calibration method designed for central omnidirectional cameras as described in the algorithm in Section 5.7. After this step, the mirror parameter (a or c), the essential matrix, and the correct point matches validated from automatically detected point correspondences are obtained.
2. Use the original point correspondences in acquired images $\{\mathbf{u'}_1 \leftrightarrow \mathbf{u'}_2\}$.
3. Set initial values. Fix $\mathtt{P}_1 = [\mathtt{I} \,|\, \mathbf{0}]$ and from the essential matrix create $\mathtt{P}_2 = [\mathtt{R}_m \,|\, \mathbf{t}_m]$. Set $\mathtt{R}_c = \mathrm{diag}(-1, 1, -1)$ (since the image plane is rotated by 180° around its y axis), $\mathbf{t}_c = (0,\, 0,\, z)^\top$, $\mathtt{K} = \begin{pmatrix} f & 0 & -u_0 \\ 0 & f & -v_0 \\ 0 & 0 & 1 \end{pmatrix}$ (zero skew and square pixels are assumed). The z is the distance between $\mathbf{F}$ and $\mathbf{C}$, the f is the focal length of the perspective camera, and $(u_0,\, v_0)^\top$ is the center of the view field ellipse. The z and the f are obtained as follows

   - PCD cameras: a is obtained by the auto-calibration. Measure z_{mm} approximately. Compute $\rho = \frac{2\,a\,K}{R}$ [pxl/mm], where K is the ratio R/H (radius/height) of the mirror. Convert z_{mm} to pixels by $z = \rho\, z_{mm}$. Compute $f = \frac{r_{max}}{4\,a\,K}(2\,z + 4\,a\,K^2 - a)$ (obtained analogically as Equation (5.18) for HCD cameras). Use the relationship $z = \frac{2\,a\,K}{r_{max}} f - 2\,a\,K^2 + \frac{a}{2}$ (obtained from the previous equation) in the optimization.
   - HCD cameras: $c = \frac{b}{a}$ is obtained by the auto-calibration. Set $a \stackrel{!}{=} 1$, $b = a\,c$ and $z = 2\,e = 2\sqrt{1 + c^2}$. Compute the focal length f from Equation (5.18). Use the relationship $z = \frac{R\,f}{r_{max}} - K\,R + e - a$ in the optimization, where R is from Equation (5.16).
   - SCD cameras: $c = \frac{d}{a}$ is obtained by the auto-calibration. Set $d \stackrel{!}{=} 1$ $a = \frac{d}{c}$. Compute $z = d + k\,a$, where $k = 0.64$, and the focal length f from Equation (5.29). Use the relationship $z = a\sqrt{\frac{f^2}{r_{max}^2} + 1}$ (obtained from Equation (5.29)) in the optimization.
4. To obtain the final 3D metric reconstruction, minimize[1] the error in Equation (6.14) over all inliers (selected by the 1st step) over a (also b for HCD cameras), $\mathtt{R}_c$, $\mathbf{t}_c$, $\mathtt{R}_m$, $\mathbf{t}_m$, $\mathtt{K}$, i.e., perform the minimization in Equation (6.15). In the minimization, the z is computed from the f, rotational matrices and the vector $\mathbf{t}_c$ are parametrized by three numbers, $\mathbf{t}_m$ by two numbers (two angles, since the magnitude is unimportant) and $\mathtt{K}$ by five numbers (two for principal point, a focal length, a skew,

[1] In our experiments, Levenberg-Marquardt minimization method was used.

and a non-square factor [43]). In the minimization process, the derived non-central camera model of a particular catadioptric camera is used.

5. After optimizing the final 3D metric reconstruction (3D points and camera positions in coordinate system of the first camera) is obtained.

6.5. Closure

In this chapter, we focused on real catadioptric cameras which become slightly non-central, e.g., due to imprecise alignment of a camera and a mirror. We derived accurate non-central and suitable approximate central models for parabolic, hyperbolic, and spherical mirrors. We designed a method allowing to build a 3D metric reconstruction from two uncalibrated non-central catadioptric images. We have shown that an accurate and very complete 3D reconstruction of surrounding scene can be obtained already from two views.

The proposed autocalibration and 3D reconstruction technique is not restricted to the three mirrors presented in the thesis. The same procedure can be applied for non-central catadioptric cameras composed of uniform resolution mirrors especially designed to fulfil a given type of projection [46, 32].

Two important conclusions follow from this chapter. First, the correspondences between catadioptric images can (and need to) be validated using an approximation of the non-central camera model by a suitable central one. Second, the non-central camera model has to be used to get geometrically correct 3D metric reconstruction. We have shown that the hierarchical approach can be used to solve the complex problem of a 3D metric reconstruction from two non-central panoramic cameras. We would like to stress once more that appropriate approximations are instrumental in estimating models from real correspondence data contaminated by outliers.

We observed that the reconstruction & auto-calibration with non-central catadioptric cameras is as easy (or as difficult) as with central catadioptric cameras, provided that the correspondence problem can be solved with a suitable approximate central model. It turns out that it is the number of parameters of the camera model that matters rather than the exact centrality of the projection. Thus, spherical catadioptric cameras become the most practical. They can be described by a simple model, and therefore easily auto-calibrated, can be easily manufactured, and provide images with fewer blur compared to other quadric mirrors.

7 Conclusion

The thesis presented the theory and practice of the two-view geometry of central omnidirectional cameras in the sense of camera auto-calibration and 3D metric reconstruction. The thesis extended a non-linear camera model estimation from epipolar geometry, introduced by Fitzgibbon in 2001, to omnidirectional cameras with the angle of view larger than 180°. It was shown that a "stronger" auto-calibration method can be obtained for omnidirectional cameras than for conventional ones, regarding both methods based on solving the Polynomial Eigenvalue Problem. In particular, we have

1. designed the robust method allowing to build a 3D *metric* reconstruction from two uncalibrated omnidirectional images from point correspondences automatically,

2. formulated a theory describing image formation and auto-calibration of central omnidirectional cameras,

3. demonstrated that real catadioptric cameras usually have to be treated as non-central cameras. It leads to a more complicated camera model. The benefit is a more precise 3D reconstruction.

It was shown that the complex problem - *Structure from Motion (SFM) from real uncalibrated panoramic cameras* - should be solved by a hierarchical approach using appropriate approximations in each step of the hierarchy. The approximations of models are instrumental to find an initial estimate of the camera model and to select correct correspondences from the automatically established tentative matches. In each further step, the more precise camera model can be used, and thus a gradual improvement of the final SFM can be achieved. We have shown, for instance, that for real (slightly non-central) catadioptric cameras the hierarchy starts with an approximate central model and ends up with the most precise non-central one. Solving the problem of SFM directly by using the most precise model is usually too complex and would make the task unsolvable.

The main advantage of our approach is that no special assumption (e.g., a calibration object or a special type of motion) has to be used to auto-calibrate a camera and to build a 3D metric reconstruction.

A Appendix

A.1. Deeper analysis of fish-eye lens functions

The model for the Nikon FC-E8 fish-eye lens from Section 5.5.2 capturing the direction from an image point radius to an angle between 3D rays and the camera optical axis is

$$\theta = \frac{a\,r}{1 + b\,r^2}\,. \tag{A.1}$$

The inverse function, i.e., from the angle to the point radius, is

$$r = \frac{a - \sqrt{a^2 - 4b\,\theta}}{2\,b\theta}\,. \tag{A.2}$$

The function $g(r = 0, a, b)$, i.e.,

$$\lim_{r\to 0} g(r, a, b) = \lim_{r\to 0} \frac{r}{\tan \frac{ar}{1+br^2}} = \frac{1}{a}\,, \tag{A.3}$$

depends on the parameter a only. The value in which $g(r, a, b)$ intersects x-axis is

$$g(r, a, b) = 0 \quad \Leftrightarrow \quad \frac{ar}{1 + br^2} = \frac{\pi}{2} \qquad \text{and} \qquad r \neq 0\,,$$

therefore

$$r_{\frac{\pi}{2}} = \frac{a - \sqrt{a^2 - \pi^2 b}}{\pi b}\,. \tag{A.4}$$

It follows from Equation (A.3) and Equation (A.4) that for small angle θ, the function $g(r, a, b)$ depends on the parameter a only and for the larger angle, the function depends on both parameters a and b. However, the parameter b affects the angle less than the a as the change of the a about 5% and the b about 100% causes the comparable error $\Delta\theta$, see Figure A.1. Therefore, the parameter a should be determined with bigger accuracy than the parameter b. Figure A.1 shows, see black dashed curve, that wrong a can be partially corrected by b and vice versa. It means that the estimation of a camera model with given tolerance $\Delta\theta$ can bias the estimated parameters a and b. However, the total error does not exceed the tolerance $\Delta\theta$.

The same result follows for the Sigma 8mm-f4-EX fish-eye lens, thus we present only the final equations. The model from Section 5.5.2 is as follows

$$\theta = \frac{1}{b} \arcsin\left(\frac{b\,r}{a}\right), \qquad\qquad r = \frac{a}{b} \sin(b\,\theta)\,. \tag{A.5}$$

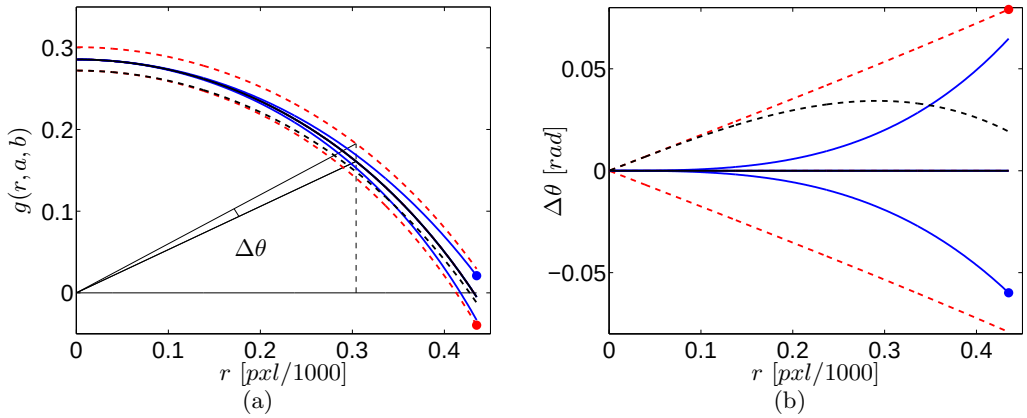


Figure A.1.: Sensitivity of the function $g(r, a, b)$ to parameters a, b for the Nikon FC-E8 fish-eye lens. (a) The graph of the function $g(r, a, b)$ w.r.t. r at a_0, b_0 (black solid curve) and for a change of the parameter $a = a_0 \pm 5\%$ (red dotted curves) and $b = b_0 \pm 100\%$ (blue solid curves). (b) The graph of the error $\Delta\theta$ caused by the change of the parameter a, resp. b. Black dashed curve depicts the total error if $a = a_0 - 5\%$ and $b = b_0 + 100\%$, i.e., the combination of the curves marked by dots at the end.

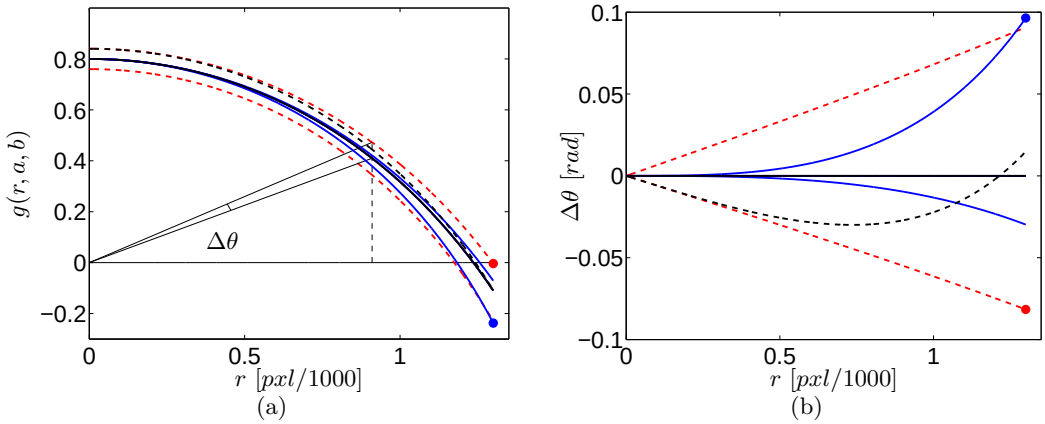


Figure A.2.: Sensitivity of the function $g(r, a, b)$ to parameters a, b for the Sigma 8mm-f4-EX fish-eye lens. See Figure A.1 for explanation.

The function $g(r = 0, a, b)$, i.e.,

$$\lim_{r\to 0} g(r, a, b) = \lim_{r\to 0} \frac{r}{\tan\left(\frac{1}{b}\arcsin\frac{br}{a}\right)} = a\,, \tag{A.6}$$

shows again that for $r = 0$ the function g depends on the a only. The value in which $g(r, a, b)$ intersects the x-axis is

$$g(r, a, b) = 0 \quad \Leftrightarrow \quad \frac{1}{b}\arcsin\frac{br}{a} = \frac{\pi}{2} \qquad \text{and} \qquad r \neq 0\,,$$

therefore

$$r_{\frac{\pi}{2}} = \frac{a}{b}\sin\left(\frac{\pi}{2}b\right)\,. \tag{A.7}$$

See Figure A.2 for results.

A.2. Solving the Polynomial Eigenvalue Problem

This section includes some remarks about solving the Polynomial Eigenvalue Problem (PEP) and its stability when dots are contaminated by Gaussian noise. The algorithms for solving the PEP can be found in [1]. MATLAB solves the PEP by the function `polyeig`.

Except for para-catadioptric cameras, all auto-calibration methods presented in the thesis lead to the PEP of degree two, i.e., to the Quadratic Eigenvalue Problem (QEP). The QEP can be easily re-arranged and solved as a generalized eigensystem [1, 30]. Let us give a quick overview how to do it.

Suppose the QEP from Equation (5.23)

$$\left(\mathtt{D}_1 + c\,\mathtt{D}_2 + c^2\,\mathtt{D}_3\right)\mathbf{f} = \mathbf{0}\,, \tag{A.8}$$

where $\mathtt{D}_i \in \mathbb{R}^{n\times n}$, $\mathbf{f} \in \mathbb{R}^n$ and define a new vector $\mathbf{e} = c\,\mathbf{f}$, giving

$$\mathtt{D}_1\mathbf{f} + c\,\mathtt{D}_2\mathbf{f} + c\,\mathtt{D}_3\mathbf{e} = \mathbf{0}\,. \tag{A.9}$$

Remind that $n = 9$ for hyperbolic and spherical mirrors and for one-parametric fish-eye models, $n = 15$ for the two-parametric fish-eye lens models. The following pair of equation can be constructed

$$\begin{aligned} -\mathtt{D}_1\mathbf{f} + 0\,\mathbf{e} &= c\,\mathtt{D}_2\mathbf{f} + c\,\mathtt{D}_3\mathbf{e}\,, \\ \mathtt{0}_{n\times n}\mathbf{f} + \mathtt{I}_{n\times n}\mathbf{e} &= c\,\mathtt{I}_{n\times n}\mathbf{f} + c\,\mathtt{0}_{n\times n}\mathbf{e}\,, \end{aligned} \tag{A.10}$$

what can be written in a matrix form as follows

$$\begin{aligned} \begin{pmatrix} -\mathtt{D}_1 & \mathtt{0} \\ \mathtt{0} & \mathtt{I} \end{pmatrix}\begin{pmatrix} \mathbf{f} \\ \mathbf{e} \end{pmatrix} &= c\begin{pmatrix} \mathtt{D}_2 & \mathtt{D}_3 \\ \mathtt{I} & \mathtt{0} \end{pmatrix}\begin{pmatrix} \mathbf{f} \\ \mathbf{e} \end{pmatrix}, \\ \mathtt{A}\mathbf{v} &= c\,\mathtt{B}\mathbf{v}\,. \end{aligned} \tag{A.11}$$

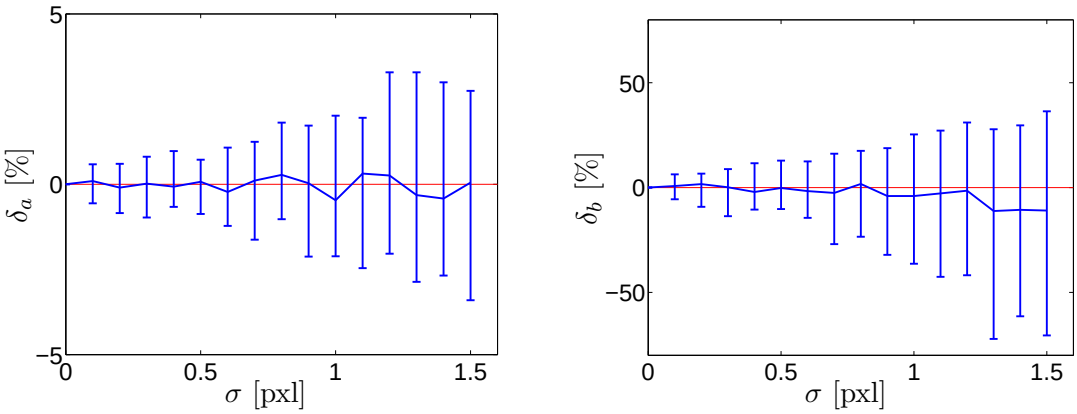


Figure A.3.: Noise influence test on the QEP based on a 15-point RANSAC with bucketing performed on synthetic data. The relative error for parameter a and the b as a function of noise level σ. The mean, 10th and 90th percentile are shown on error bars.

Equation (A.11) can be solved as the generalized eigensystem or it can be converted to the regular eigensystem iff the $\mathtt{D}_1$ is non-singular as

$$\frac{1}{c}\mathbf{v} = \mathtt{C}\,\mathbf{v}, \qquad \text{where} \quad \mathtt{C} = \begin{pmatrix} -\mathtt{D}_1^{-1} & \mathtt{0} \\ \mathtt{0} & \mathtt{I} \end{pmatrix} \begin{pmatrix} \mathtt{D}_2 & \mathtt{D}_3 \\ \mathtt{I} & \mathtt{0} \end{pmatrix},$$

and solved by standard methods [1, 41]. Equation (A.8) and Equation (A.11) give $2n$ solutions for c and $\mathbf{f}$. Fitzgibbon showed in [30] that only $n + \mathrm{rank}(\mathtt{D}_3)$ solutions are finite, from which some are imaginary, and therefore, in practice, only 1-4 solutions are considerable.

Noise

Let us investigate the behaviour of the QEP based on the RANSAC with bucketing on synthetic data corrupted by Gaussian noise. A synthetic scene was generated using the following procedure. First, random points were generated in the first image. Then, the Sigma fish-eye lens model, Equation (5.33), was used with the ground truth parameters $a = 0.8$ and $b = 0.2$ (corresponding to Canon EOS–1Ds digital camera) to obtain 3D rays going from the optical center to the scene. Next, 3D scene points were established by randomly varying depths along the rays. Finally, we set a rotation and a translation of the second camera w.r.t. the first one and projected the 3D points to the second image. By this, we obtained the set of correspondences $\mathbf{u}_1 \leftrightarrow \mathbf{u}_2$.

To simulate a real situation, we added Gaussian noise with zero mean and standard deviation σ to each point correspondence in both images. We applied the algorithm, described in Section 5.7, used in three level hierarchy finishing with 15-point RANSAC described in Section 5.5.4. On each noise level with a constant σ, 50 iterations were performed. The relative error of the a and b w.r.t. the noise level σ is shown in Figure A.3.

It can be seen that increasing the noise level increases the interval about the median value. Notice that δ_b is 10 times larger that δ_a. It is the consequence of the result achieved in the Section A.1 that the parameter b has lower influence on the model accuracy than the parameter a.

A.3. Derivation of some formulae in Maple

This section shows how MAPLE [49] was used to obtain some analytical forms derived in the thesis.

Partial derivatives

The partial derivatives $\frac{\partial p_i}{\partial c}$ from Equation (5.21) were obtained by

```
> dpi_dc := taylor(pi,c=c0,2);
```

where `pi` stands for the i-th coordinate (i={x, y, z}) of vector $\mathbf{p}$, Equation (5.19), for HCD cameras, and Equation (5.28) for SCD cameras.

PEP matrices

The matrices $\mathtt{D}_i$ in Equation (5.24) for the PEP in Equation (5.23) for HCD cameras are obtained as follows

```
> F  := matrix([[f1,f2,f3],[f4,f5,f6],[f7,f8,f9]]);
> XX := matrix([[pp], [qq], [ww]]);
> SS := matrix([[ss], [tt], [oo]]);
> X  := matrix([[p], [q], [w]]);
> S  := matrix([[s],[t],[o]]);
> f  := evalm(transpose((XX+b*SS))&*F&*(X+b*S));
> collect(f[1,1],b);
```

The result obtained after running the last command is the polynomial in b, which can easily be arranged into the matrices $\mathtt{D}_i$ to get the PEP.

Non-central model

The polynomial giving the solution λ for non-central PCD cameras, Equation (6.5), can be obtained by the following

```
> Q := matrix([[-1/a^2,0,0,0],[0,-1/a^2,0,0],
              [0,0,0,-1/a],[0,0,-1/a,1]]);
> v := matrix([[lam*r],[lam*s],[lam*t],[1]]);
> t := matrix([[k],[l],[m],[0]]);
> f := evalm(transpose(v+t)&*Q&*(v+t));
```

```
> collect(f[1,1],lam);
```

Similar commands were used in other derivations of the equations in Chapter 6.

A.4. Directories and m-files

This section is relevant for internal users of CMP (Center for Machine Perception) network only. The m-files should help to continue in our work. All useful m-files related to the thesis are included in the following directories

```
/home.dokt/micusb1/Work/Soft/Fishcal/
/home.dokt/micusb1/Work/Soft/Mircal/
/home.dokt/micusb1/matlab/fisheye/
```

The following m-files implement the auto-calibration and the 3D reconstruction, respectively:

- fish-eye lenses:

  ```
  /home.dokt/micusb1/Work/Soft/Fishcal/fepseq_nik5b.m
  /home.dokt/micusb1/Work/Soft/Fishcal/reconst.m
  ```

- non-central PCD cameras:

  ```
  /home.dokt/micusb1/Work/Soft/Fishcal/fepseq_nik5b.m
  /home.dokt/micusb1/Work/Soft/Mircal/recons_nc_pmir.m
  ```

- non-central HCD cameras:

  ```
  /home.dokt/micusb1/Work/Soft/Mircal/Hmir/hm_calib_ep.m
  /home.dokt/micusb1/Work/Soft/Mircal/recons_nc_hmir.m
  ```

- non-central SCD cameras:

  ```
  /home.dokt/micusb1/Work/Soft/Mircal/Smir/sm_calib_ep.m
  /home.dokt/micusb1/Work/Soft/Mircal/reconst_nc_smir.m
  ```

Bibliography

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, editors. *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. SIAM, Philadelphia, 2000.

[2] S. Baker and S. K. Nayar. A theory of catadioptric image formation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 35–42. Narosa Publishing House, January 1998.

[3] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision (IJCV)*, 35(2):175–196, 1999.

[4] H. Bakstein and T. Pajdla. Panoramic mosaicing with a 180° field of view lens. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 60–67. IEEE Computer Society Press, June 2002.

[5] J. P. Barreto and H. Araujo. Issues on the geometry of central catadioptric image formation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages II:422–427. IEEE Computer Society Press, December 2001.

[6] J. P. Barreto and H. Araujo. Geometric properties of central catadioptric line images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 2353 of *Lecture Notes in Computer Science*, pages 237–251, May 2002.

[7] J. P. Barreto and H. Araujo. Direct least square fitting of paracatadioptric line images. In *Proceedings of the IEEE Workshop on Omnidirectional Vision and Sensor Networks*. IEEE Computer Society Press, June 2003.

[8] E. Bayro-Corochano and C. Lopez-Franco. Omnidirectional vision: Unified model using conformal geometry. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3021 of *Lecture Notes in Computer Science*, pages 536–548. Springer Verlag, May 2004.

[9] S. S. Beauchemin, R. Bajcsy, and G. Givaty. A unified procedure for calibrating intrinsic parameters of fish-eye lenses. In *Proceedings of the Conference on Vision Interface*, pages 272–279, May 1999.

[10] R. Benosman and S. B. Kang, editors. *Panoramic vision: sensors, theory, and applications*. Monographs in computer science. Springer Verlag, New York, 2001.

[11] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 591–596. IEEE Computer Society Press, October 2003.

[12] 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 Conference on Image Processing (ICIP)*, pages 225–228. IEEE Computer Society Press, October 2001.

[13] T. Brodsky, C. Fermuller, and Y. Aloimonos. Directions of motion fields are hardly ever ambiguous. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1064 of *Lecture Notes in Computer Science*, pages 119–128, April 1996.

[14] T. Brodsky, C. Fermuller, and Y. Aloimonos. Directions of motion fields are hardly ever ambiguous. *International Journal of Computer Vision (IJCV)*, 26(1):5–24, 1998.

[15] A. M. Bruckstein and T. J. Richardson. Omniview cameras with curved surface mirrors. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 79–84. IEEE Computer Society Press, June 2000.

[16] R. Bunschoten and B. Kröse. Robust scene reconstruction from an omnidirectional vision system. *IEEE Transaction on Robotics and Automation*, 19(2), 2002.

[17] P. Chang and M. Hebert. Omni-directional structure from motion. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 127 – 133. IEEE Computer Society Press, June 2000.

[18] O. Chum, J. Matas, and Š. Obdržálek. Enhancing ransac by generalized model optimization. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, volume II, pages 812–817, January 2004.

[19] O. Chum, T. Werner, and T. Pajdla. Joint orientation of epipoles. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, pages 73–82. The British Machine Vision Association, September 2003.

[20] S. Derrien and K. Konolige. Approximating a single viewpoint in panoramic imaging devices. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 85–90. IEEE Computer Society Press, June 2000.

[21] R. Descartes and D. Smith. *The geometry of René Descartes*. Dover Publ.: New York. Originally published in Discours de la Methode, 1637.

[22] F. Devernay and O. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *The International Society for Optical Engineering (SPIE) proceedings - Conference on Investigate and Trial Image Processing*, volume 2567, pages 62–72, July 1995.

[23] P. Doubek and T. Svoboda. Reliable 3D reconstruction from a few catadioptric images. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 71–78. IEEE Computer Society Press, June 2002.

[24] J. Fabrizio, J. P. Tarel, and R. Benosman. Calibration of panoramic catadioptric sensors made easier. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 45–52. IEEE Computer Society Press, June 2002.

[25] H. Farid and A. C. Popescu. Blind removal of image non-linearities. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 1, pages 76–81. IEEE Computer Society Press, July 2001.

[26] O. Faugeras, Q. Tuan Luong, and T. Papadopoulo. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications.* MIT Press, Cambridge, Massachusetts, 2001.

[27] C. Fermuller and Y. Aloimonos. Ambiguity in structure from motion: Sphere vs. plane. *International Journal of Computer Vision (IJCV)*, 28:137–154, 1998.

[28] R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume 1. Addison-Wesley, 1963.

[29] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the Association for Computing Machinery (CACM)*, 24:381, June 1981.

[30] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 125–132. IEEE Computer Society Press, December 2001.

[31] M. M. Fleck. Perspective projection: the wrong imaging model. Technical Report TR 95-01, Department of Computer Science, University of Iowa, 1995.

[32] S. Gächter, T. Pajdla, and B. Mičušík. Mirror design for an omnidirectional camera with a space variant imager. In *Proceedings of the IEEE Workshop on Omnidirectional Vision Applied to Robotic Orientation and Nondestructive Testing (NDT)*, pages 99–105. IEEE Computer Society Press, August 2001.

[33] J. Gaspar, C. Decco, J. Okamoto, Jr., and J. Santos-Victor. Constant resolution omnidirectional cameras. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 27–34. IEEE Computer Society Press, June 2002.

[34] C. Geyer and K. Daniilidis. Catadioptric camera calibration. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 398–404. IEEE Computer Society Press, September 1999.

[35] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical applications. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1842 of *Lecture Notes in Computer Science*, pages 445–461. Springer Verlag, June 2000.

[36] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 279–286. IEEE Computer Society Press, December 2001.

[37] C. Geyer and K. Daniilidis. Para-catadioptric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(5):687–695, May 2002.

[38] C. Geyer and K. Daniilidis. Conformal rectification of an omnidirectional stereo pair. In *Proceedings of the IEEE Workshop on Omnidirectional Vision and Sensor Networks*. IEEE Computer Society Press, June 2003.

[39] C. Geyer and K. Daniilidis. Mirrors in motion: Epipolar geometry and motion estimation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 766–773. IEEE Computer Society Press, October 2003.

[40] J. Gluckman and S. K. Nayar. Ego-motion and omnidirectional cameras. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 999–1005. Narosa Publishing House, January 1998.

[41] G. H. Golub and C. F. Van Loan. *Matrix Computation*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, USA, 3rd edition, 1996.

[42] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 108–115. IEEE Computer Society Press, July 2001.

[43] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[44] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding (CVIU)*, 68(2):146–157, 1997.

[45] E. Hecht and A. Zajac. *Optics*. Addison Wesley, 1974.

[46] A. Hicks and R. Bajcsy. Catadioptric sensors that approximate wide-angle perspective projections. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 97–103. IEEE Computer Society Press, June 2000.

[47] R. Hicks and R. Perline. Equi-areal catadioptric sensors. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 13–18. IEEE Computer Society Press, June 2002.

[48] `http://www.2d3.com`.

[49] `http://www.maplesoft.com`.

[50] `http://www.meopta.cz`.

[51] `http://www.neovision.cz`.

[52] `http://www.remotereality.com`.

[53] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(2):257–262, 1992.

[54] S. B. Kang. Catadioptric self-calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 201–207. IEEE Computer Society Press, June 2000.

[55] S. B. Kang and R. Szeliski. 3-D scene data recovery using omnidirectional multibaseline stereo. *International Journal of Computer Vision (IJCV)*, 25(2), November 1997.

[56] J. Kumler and M. Bauer. Fish-eye lens designs and their relative performance. In *The International Society for Optical Engineering (SPIE) proceedings - Current Developments in Lens Design and Optical Systems Engineering*, volume 4093, pages 360–369, October 2000.

[57] D. Martinec and T. Pajdla. Structure from many perspective images with occlusions. In *Proceedings of European Conference on Computer Vision (ECCV)*, volume II, pages 355–369. Springer-Verlag, May 2002.

[58] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, pages 384–393. The British Machine Vision Association, September 2002.

[59] J. Mellor. Geometry and texture from thousands of images. *International Journal of Computer Vision (IJCV)*, 51(1), 2003.

[60] B. Mičušík, D. Martinec, and T. Pajdla. 3D metric reconstruction from uncalibrated omnidirectional images. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, volume I, pages 545–550. Asian Federation of Computer Vision Societies, January 2004.

[61] B. Mičušík and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. Research Report CTU–CMP–2002–12, Center for Machine Perception, K333 FEE Czech Technical University, Prague, Czech Republic, June 2002.

[62] B. Mičušík and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 485–490. IEEE Computer Society Press, June 2003.

[63] B. Mičušík and T. Pajdla. Omnidirectional camera model and epipolar geometry estimation by RANSAC with bucketing. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, number 2749 in Lecture Notes in Computer Science, pages 83–90. Springer Verlag, June – July 2003.

[64] B. Mičušík and T. Pajdla. Using RANSAC for omnidirectional camera model fitting. In *Proceeding of the 8th Computer Vision Winter Workshop*, pages 153–158. Czech Pattern Recognition Society, February 2003.

[65] B. Mičušík and T. Pajdla. Autocalibration & 3D reconstruction with non-central catadioptric cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, June – July 2004.

[66] B. Mičušík and T. Pajdla. Para-catadioptric camera auto-calibration from epipolar geometry. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, volume II, pages 748–753. Asian Federation of Computer Vision Societies, January 2004.

[67] S. K. Nayar. Catadioptric omnidirectional camera. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 482–488. IEEE Computer Society Press, June 1997.

[68] S. A. Nene and S. K. Nayar. Stereo with mirrors. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume II, pages 1087–1094. Narosa Publishing House, January 1998.

[69] J. Neumann, C. Fermüller, and Y. Aloimonos. Eye design in the plenoptic space of light rays. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1160–1167. IEEE Computer Society Press, October 2003.

[70] J. Oliensis. Exact two–image structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(12):1618–1633, 2002.

[71] T. Pajdla. Epipolar geometry of some non-classical cameras. In *Proceedings of the Computer Vision Winter Workshop*, pages 223–233. Slovenian Pattern Recorgnition Society, February 2001.

[72] T. Pajdla. *Stereo Geometry of Non-Central Cameras*. PhD thesis, Czech Technical University in Prague, March 2003.

[73] T. Pajdla, T. Svoboda, and V. Hlaváč. Epipolar geometry of central panoramic cameras. In R. Benosman and S. B. Kang, editors, *Panoramic Vision: Sensors, Theory, and Applications*, pages 85–114. Springer Verlag, Berlin, Germany, 1 edition, 2001.

[74] T. Pajdla, T. Werner, and V. Hlaváč. Correcting radial lens distortion without knowledge of 3-D structure. Technical Report K335-CMP-1997-138, Czech Technical University, Prague, Czech Republic, 1997.

[75] R. Pless. Using many cameras as one. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 587–593. IEEE Computer Society Press, June 2003.

[76] S. F. Ray. *Applied photographic optics : Lenses and optical systems for photography, film, video, electronic and digital imaging*. Focal Press, Oxford, 3rd edition, 2002.

[77] S. Salvarese, M. Chen, and P. Perona. Recovering local shape of a mirror surface from reflection of a regular grid. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3023 of *Lecture Notes in Computer Science*, pages 468–481. Springer Verlag, May 2004.

[78] S. Shah and J. K. Aggarwal. Intrinsic parameter calibration procedure for a (high distortion) fish-eye lens camera with distortion model and accuracy estimation. *Pattern Recognition*, 29(11):1775–1788, 1996.

[79] J. Šivic. Geometry of concentric multiperspective panoramas. Master's thesis, Center for Machine Perception, K333 FEE Czech Technical University, Prague, Czech Republic, January 2002.

[80] L. Spacek. Coaxial omnidirectional stereopsis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3021 of *Lecture Notes in Computer Science*, pages 354–365. Springer Verlag, May 2004.

[81] G. P. Stein. Lens distortion calibrating using point correspondences. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 602–609. IEEE Computer Society Press, June 1997.

[82] D. Strelow, J. Mishler, D. Koes, and S. S. Precise omnidirectional camera calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume I, pages 689–694. IEEE Computer Society Press, December 2001.

[83] P. Sturm. A method for 3D reconstruction of piecewise planar objects from single panoramic images. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 119–126. IEEE Computer Society Press, June 2000.

[84] P. Sturm. Mixing catadioptric and perspective cameras. In *Proceedings of the IEEE Workshop on Omnidirectional Vision*, pages 60–67. IEEE Computer Society Press, June 2002.

[85] T. Svoboda and T. Pajdla. Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision (IJCV)*, 49(1):23–37, August 2002.

[86] T. Svoboda, T. Pajdla, and V. Hlaváč. Epipolar geometry for panoramic cameras. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1406 of *Lecture Notes in Computer Science*, pages 218–232. Springer Verlag, June 1998.

[87] T. Svoboda, T. Pajdla, and V. Hlaváč. Motion estimation using central panoramic cameras. In *In Proceeedings of the IEEE International Conference on Intelligent Vehicles*, pages 335–340. Causal Productions, October 1998.

[88] R. Swaminathan, M. D. Grossberg, and S. K. Nayar. Caustics of catadioptric cameras. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume II, pages 2–9. IEEE Computer Society Press, July 2001.

[89] R. Swaminathan, M. D. Grossberg, and S. K. Nayar. Designing mirrors for catadioptric systems that minimize image errors. In *Proceedings of the Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras*. CD-ROM edition, May 2004.

[90] R. Swaminathan and S. K. Nayar. Non-metric calibration of wide-angle lenses and polycameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(10):1172–1178, 2000.

[91] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, 43(2):235–286, 2001.

[92] T. Werner. Combinatorial constraints on multiple projections of a set of points. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume II, pages 1011–1016. IEEE Computer Society Press, October 2003.

[93] T. Werner. Constraint on five points in two images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume II, pages 203–208. IEEE Computer Society Press, June 2003.

[94] T. Werner and T. Pajdla. Cheirality in epipolar geometry. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 548–553. IEEE Computer Society Press, July 2001.

[95] T. Werner, T. Pajdla, and M. Urban. Practice of 3D reconstruction from multiple uncalibrated unorganized images. In *Proceedings of the Czech Pattern Recognition Workshop*, pages 71–76. Czech Pattern Recognition Society, February 2000.

[96] Y. Xiong and K. Turkowski. Creating image-based VR using a self-calibrating fisheye lens. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 237–243. IEEE Computer Society Press, June 1997.

[97] X. Ying and Z. Hu. Catadioptric camera calibration using geometric invariants. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1351–1358. IEEE Computer Society Press, October 2003.

[98] X. Ying and Z. Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 3021 of *Lecture Notes in Computer Science*, pages 442–355. Springer Verlag, May 2004.

[99] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 407–411. IEEE Computer Society Press, August 1996.

[100] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence (AI)*, 78(1-2):87–119, 1995.