



HAL
open science

Image Geometry

Peter Sturm

► **To cite this version:**

Peter Sturm. Image Geometry. Pascal Vasseur; Fabio Morbidi. Omnidirectional Vision: From Theory to Applications, ISTE, 2022. hal-03563140v1

HAL Id: hal-03563140

<https://inria.hal.science/hal-03563140v1>

Submitted on 9 Feb 2022 (v1), last revised 11 Feb 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Image Geometry

Peter Sturm

peter.sturm@inria.fr

Initial version of a chapter in “Omnidirectional Vision: From Theory to Applications”, edited by Pascal Vasseur and Fabio Morbidi, to appear in 2022

1 Introduction

Geometry is important in various aspects of omnidirectional vision, from sensor and camera design and modeling, over image analysis, to structure from motion. The probably most fundamental issue concerns design and modeling: how to build image acquisition devices that have certain desired characteristics? Foremost among these is of course the desire to acquire images with a very wide field of view, be it panoramic, hemispheric, fully spherical, or somewhere in between¹. The potential interests are clear – such a wide field of view allows to visualize or analyze a scene more completely² and for example, to detect obstacles or objects to interact with, all around a robot. It also turns out that ego-motion estimation is generally more stable and accurate with a large field of view [27]. Besides such practical interests, there may also be others, such as esthetics.

Various technical solutions have been developed to acquire omnidirectional images. The first ones relied on rotating a regular or tailor-made camera about itself and “stitching together” images acquired during the rotation such as to form an image representing an extended field of view, an approach nowadays provided as a basic feature on most consumer-grade digital cameras. The obvious disadvantages are that image acquisition is not instantaneous, making it difficult to operate in dynamic contexts or to use it for omnidirectional video acquisition, and that it requires image processing that may not succeed for all types of scenes. Alternative solutions were thus developed, especially of two types. One consists in developing lens designs capable of delivering the sought after large fields of view, in particular fisheye objectives with fields of view nowadays even exceeding 180°. The other, quite popular in robotics in the last decades, is to use mirrors to enhance the field of view of a camera, leading to systems baptized catadioptric cameras³.

Besides an extended field of view, various other design objectives have been pursued by scientists and engineers. Among the most important is the aim of achieving a single effective viewpoint or optical center – in the following we will also speak of *central cameras*. An example is given in figure 1.

Why is this property interesting? First, as noted by [1], images acquired by central cameras allow to synthesize perspectively correct images, i.e. images as if acquired from the same viewpoint by a perspective camera. This property has for instance be used to “navigate” in panoramic images through panning one’s virtual viewing direction and displaying a perspective image generated from

¹In the following, for brevity, we will mostly use the term omnidirectional when speaking of very large fields of view.

²Among the first applications were the creation of images for panoramic viewing and the study of cloud formations from hemispheric images of the sky. Surveillance type applications are of course also a natural client.

³Catadioptric meaning the combination of both refractive and reflective optical elements (lenses and mirrors).

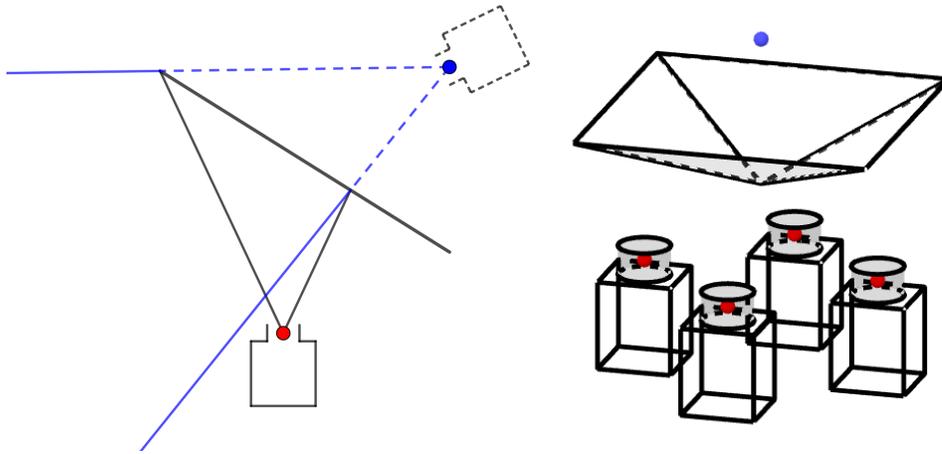


Figure 1: Left: a camera pointed at a planar mirror acquires the same image (besides issues such as loss of sharpness or color richness due to imperfect reflection) as would a virtual camera situated and oriented symmetrically on the other side of the mirror.

Right: several camera–mirror pairs are arranged such that the corresponding virtual cameras have the same optical centers. This arrangement allows to produce a panoramic image as if it were taken from a single effective viewpoint inside the pyramid of mirrors – it thus corresponds to a central camera [20, 24]. To get a complete panoramic field of view, an adequate number of camera–mirror pairs must be used, depending on the camera’s individual fields of view.

the panorama according to the current viewing direction [7]. Obviously, this idea is applicable for any combination of real and virtual camera, provided they are both central and that the real camera offers a sufficient field of view to synthesize images for the virtual camera. Examples are shown in the accompanying videos `Perspective_Rendering.mp4` and `Panoramic_Rendering.mp4`, where it is illustrated how to synthesize perspective respectively panoramic images, from an image acquired with an omnidirectional catadioptric camera.

Another advantage of having a single effective viewpoint is that this opens the way to directly applying the rich toolkit of Structure-from-Motion (SfM) methods originally developed for perspective cameras. A final aspect mentioned here is that usage of central cameras enables to perform dense stereovision very efficiently, in a manner analogous to that originally developed for perspective images, through a generalized rectification process followed by “scan-line matching”. All these aspects are further developed later in this chapter.

Other design objectives besides large field of view and single effective viewpoint concern for instance the types of distortion inevitably generated by omnidirectional cameras. Some cameras are built in order to conform to an equi-angular (sometimes also called equi-distant) distortion profile: this is the case if the angle spanned by the optical axis and the viewing direction associated with a pixel in the image is proportional to the pixel’s distance from the distortion center in the image. Other distortion profiles may be interesting such as ones corresponding to area-preserving projections [18]. The choice of distortion profile or other design feature may depend on the practical application of a camera, for instance the requirement that certain parts of the scene appear in higher resolution than others. A general framework for specifying such properties and developing a dedicated mirror shape achieving these as best as possible, was proposed by [39, 17].

Various other design objectives have of course been explored, such as on optical properties (ease of focusing, color fidelity, etc.) or the volume of a camera: to circumvent the classical disadvantage of catadioptric cameras concerning their bulk, “folded” catadioptric systems composed of several nested mirrors were proposed [25].

Let us close this discussion by mentioning that while a single effective viewpoint is an attractive property, as explained above, designing cameras that are non-central on purpose may be beneficial in other ways: this gives obviously more degrees of freedom to optimize certain properties and also, non-central cameras may allow to determine the absolute scale of ego-motion and 3D reconstruction, which is not possible with central cameras. Non-central cameras are further discussed in section 4.

In this chapter, we discuss geometrical aspects of omnidirectional cameras. When discussing “the geometry” of a certain type of camera, one usually refers to some model of the image formation process: how a camera generates an image of the scene it is pointed at. The actual image formation process carried out by a real camera, is relatively complex; light emitted by objects in the scene travels through space and, if entering the camera aperture, travels across the camera’s optical elements (lenses, mirrors) and finally hits the optical sensor, producing an electrical charge (for digital cameras) that is eventually converted into greylevels or colors, for displaying an image to a human or for processing it by a computer. Most models of this process used in computer vision, are simplified representations of it. In this chapter, we are concerned with purely geometric models, whose main basic operation is to determine, given the location of a point-like object in the scene, where the object’s image will be observed on the sensor. All higher level geometric operations in computer vision, such as computing the motion of a camera from two or more images of a scene, are derived from this basic operation.

Outline of this chapter. In the next section, we briefly outline the general image formation process and explain in which ways the usual camera models, even the classical perspective or pinhole model, are a simplification thereof. Two fundamental concepts, projection and back-projection, are very briefly introduced in the subsequent section, followed by a short discussion on central and non-central cameras. When talking about 3D geometry for computer vision, a useful distinction is between what happens “outside” cameras from their “inner workings”. By “outside”, we essentially mean the geometric relations between the 3D scene and one or more cameras and more particularly, the question how to represent and infer information on the relative positioning between cameras and/or entities in the scene. This is the subject of a section, where it is addressed, for conciseness, for the case of fully calibrated cameras. The subsequent two sections then consider the geometry of the inner workings of cameras and the epipolar geometry of a pair of cameras, with an attention to the subject of rectification for stereo matching.

This chapter does not contain equations, only geometry – algebraic formulations of some of the material covered can be found in other chapters of this book and another good starting point may be [38].

2 Image Formation and Point-Wise Approximation

The generation of a digital image of a scene, by means of a camera, is a complex process. It is ultimately created through photons that cause an electric response in the camera sensors, be they made of CCD, CMOS or another technology. We may distinguish what happens inside a camera, to photons that enter its aperture, from what is going on outside: these photons result from a potentially infinitely complex game, being emitted, reflected, refracted etc. from objects in the scene, where light is “bounced” repeatedly from one object to another. Not to speak of phenomena such as atmospheric or submarine diffractions caused by particles suspended in the air or water.

All these aspects of the image formation process have been studied, in various degrees, in photogrammetry, computer vision, and computer graphics, guided by the motivation to synthesize as realistic

images as possible, to enable a meaningful analysis of imagery acquired in bad weather or otherwise “unusual” conditions such as under water, or even to “look around the corner” [41], i.e. to use an acquired image to infer something about an object hidden from the camera’s field of view through its reflection or shadows produced on other objects.

Most works in computer vision rely on different levels of approximation of the image formation process. A first such approximation is to ignore the multiple bounces light undergoes in the scene: one implicitly assumes that each point on the surface of an object in the scene emits/reflects light, and only photons reaching the camera on a direct path, are considered. Sometimes, the “light emitted” by a point is modeled by a “color” associated to that point or, more generally, by a reflectance model which represents how light impinging on that point from a direct light source, is reflected in different directions. In any such case, whenever the camera aperture is of finite extent (i.e. is not assumed to be a single point), the camera captures an entire volume of light emitted/reflected by any individual scene point. Unless the camera optics are perfect and the point is perfectly in focus, this set of light rays hits the camera sensor on a finite area, i.e. the “image” of the scene point is not confined to an infinitesimal image point. This is of course one of the well-known causes of blur. Image formation models that mimic this, have been proposed in computer vision and graphics, such as for example through the definition of point spread functions⁴.

A second level of approximation comes essentially down to ignoring blur and to assuming that light emitted by a single scene point, manifests itself in a single point in the final image. Besides the above sketched cause for blur, this approximation also ignores that in a real digital camera, light is captured through a finite set of photosensitive elements, each one capturing light within a finite area⁵.

Most works tagged as “geometry” in computer vision, use such an approximation: the image of a scene point, is again a point. Camera models are then conceived that essentially allow to compute, given the location and orientation of a camera and the location of a scene point, the location of the image point associated with that scene point. Such camera models are the basis for various tasks such as estimating the location and orientation of an object relative to a camera (pose estimation), estimating the motion of a camera just by analyzing different images taken during that motion, estimating a 3D model of the scene, etc. We will come back to these tasks in later sections. Before doing so, we first investigate, in the next section, camera models and how they represent the mapping of a scene point to an image point.

3 Projection and Back-Projection

The simplest and most widely used camera model in computer vision is the so-called pinhole model, performing a perspective projection⁶. It consists of two elements: the entire optics is represented by a single point, the so-called optical center, and the image sensor is represented by a mathematical plane, the image plane. The basic operation, the **projection** of a scene point to the image, is performed as follows: first one creates the mathematical line that connects the optical center and the scene point and second, one determines the point where that line intersects the image plane – the image point.

These simple geometrical operations can be expressed in a similarly simple algebraic manner (see for instance [38]).

As explained in section 2, this model represents of course multiple approximations to the functioning

⁴https://en.wikipedia.org/wiki/Point_spread_function

⁵We also ignore aspects such as blooming across different photosensitive elements, black current, etc.

⁶Even simpler are the affine approximations thereof, such as the orthographic model.

of a real camera. Even so, it has been observed that it models “regular” cameras (e.g. consumer cameras) sufficiently well, i.e. when being employed for tasks such as 3D reconstruction and motion estimation, the results are of acceptable accuracy. The approximation becomes insufficient though when radial or other distortions become noticeable, for instance in wide field of view cameras, or whenever the maximum possible accuracy is sought after. This can be taken care of by extending the perspective camera model accordingly, as has been studied in photogrammetry for instance since more than a century ago, resulting in more complex camera models (both geometrically and algebraically). However, even these classical extensions of the perspective model are not sufficient when considering omnidirectional cameras; we will consider these further below.

Let us now study an important concept – the reciprocal operation to projection, which we call **back-projection**. Back-projection starts from an image point and tries to answer the question where the original scene point could possibly be located. In general, unless additional information is available, the answer corresponds to a (half-) line. Back-projection for the perspective model is straightforward: one determines the line connecting the image point and the optical center, followed by “clipping” it to a half-line, for instance at the optical center or at some minimal viewing distance in front of it. Both operations have a simple algebraic expression, much like for projection.

We now move towards the case of a catadioptric camera – a system composed of a perspective camera and a mirror into which this camera gazes⁷. Suppose that the entire geometry of the system is known: position of the image plane and the optical center, as well as the shape and position of the mirror. Let us first consider the case of a general mirror shape, i.e. it is not constrained to be a surface of revolution or otherwise specific. Suppose that the mirror shape is expressed by a scalar function that takes 3D point coordinates as argument and returns the distance of the 3D point to the closest point on the mirror surface. Further suppose that we have a function that maps 3D points lying on the mirror surface, to the associated mirror’s tangent plane⁸.

It is probably obvious that in the general case, neither projection nor back-projection have a closed-form solution. Let us now show that back-projection is in general simpler to perform than projection. Back-projection can be solved as follows. Start from a point in the perspective camera’s image plane. We back-project from the perspective camera, as shown above, which results in a half-line determined by the image point and the optical center. To determine where the original 3D point could be located, we need to “back-track” the projection path. Here, we first have to determine where our half-line intersects the mirror. It should be clear that this can be done by some one-dimensional search procedure (search along the half-line). Once this intersection point is known, we retrieve the mirror’s tangent plane at that point and reflect the half-line in it; the extension of the resulting line segment to a half-line constitutes the final result of the back-projection problem.

As for the *projection* problem, it turns out to be more complex: we need to determine a point on the mirror surface, such that the reflection in the associated tangent plane, of the line spanned by that point and the 3D scene point, is a line that contains the perspective camera’s optical center. Once this is known, it merely remains to compute the final image point by intersecting that reflected line with the image plane. In general, the first step here, the determination of the appropriate point on the mirror surface, comes down to a two-dimensional search. Hence, projection is more difficult than back-projection in this general catadioptric setting.

A few comments can be made. First, if the mirror shape is a surface of revolution and if the perspective optical center lies on the axis of revolution, then projection can be reduced to a one-dimensional

⁷Note that the camera gazing at the mirror may be non-perspective, for instance a regular camera with radial distortions. But most works on catadioptric systems assume a perspective camera and for ease of exposition, we stick to this convention in the following.

⁸The following observations also apply when using other representations of the mirror shape.

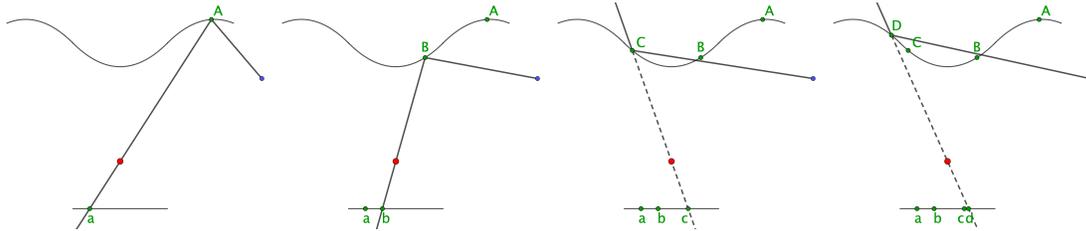


Figure 2: A camera, composed of an optical center – in red – and an image plane, looks at a curved mirror (composed of three parabolic arcs). The goal is to find the possible image points of the scene point (in blue). The four possible image points are shown, in each case with the corresponding point on the mirror where the reflection takes place. Point a and b are physically correct image points, whereas c and d are spurious mathematical solutions (they lie on the infinite lines resulting from the reflection of the scene point in mirror points C and D). The accompanying video `Three_Parabolae.mp4` shows an animation for this scenario.

search, like back-projection. Second, projection can actually lead to multiple solutions, at least if the mirror is not convex (relative to the “side” that is visible to the scene). This is shown in figure 2. It shows a two-dimensional example where a point is actually projected to multiple image points. Two of these correspond to physically valid solutions, the others are only mathematical solutions but which cannot be realized in practice. Note however that in special cases, such as for central catadioptric cameras (see below), where closed-form algebraic formulations for projection exist, these generally deliver all mathematically valid solutions, not only the physically realizable ones. As for back-projection, there is only a single physically valid solution⁹. Additional *mathematical* solutions may still exist though, like for projection. A third comment is that, as already hinted on just above, certain mirror shapes enable a closed-form algebraic expression, especially for back-projection but possibly also for projection.

Figure 3 shows the same situation with an elliptical mirror. While there is only a single physically realizable image point exists in this case, there are in total four mathematical solutions. If the optical center of the camera is situated at one of the ellipse’s focii, then the entire system becomes central (see next section) and also, the four mathematical solutions collapse to two.

⁹Besides in cases where the mirror surface is allowed to be discontinuous.

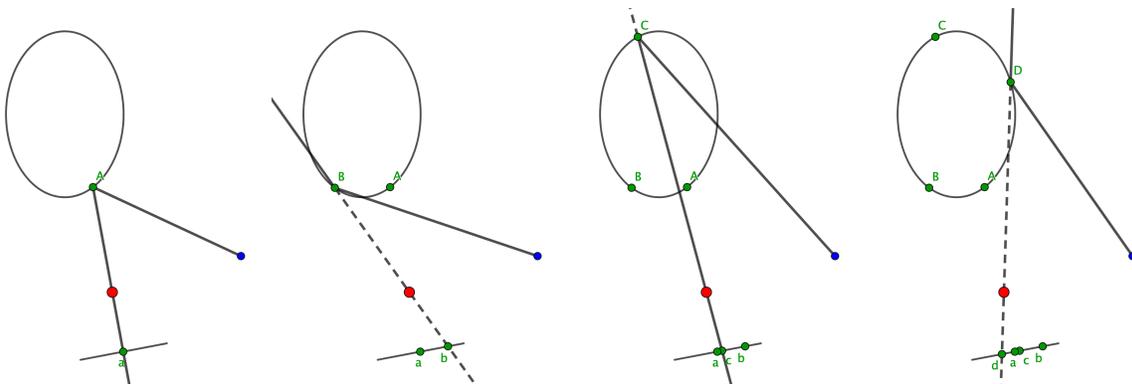


Figure 3: Same situation as in figure 2, but with an elliptical mirror. Only image point a is physically realizable, the others are spurious mathematical solutions. See the video `Ellipse_Projection.mp4` for an animation.

4 Central and non-central cameras

The notion of central camera has already been defined in section 1. Using the concept of back-projection described in section 3, we can define central cameras as those where all back-projection lines are incident with a point, the camera's (effective) optical center or viewpoint. Cameras not satisfying this condition are termed non-central. Special cases exist, such as cameras where all back-projection rays are incident with a line or line segment – in [34] these were called axial cameras.

In the following, we consider these notions in particular for catadioptric cameras. First examples of central catadioptric systems are already given in section 1, cf. figure 1. The most trivial setup consists of a central camera looking into a planar mirror. And, as shown above, arranging several camera–mirror pairs appropriately leads to a central image acquisition device with a compound field of view that is fully panoramic.

The obvious main drawback of this design is the requirement of multiple cameras. How to obtain an omnidirectional field of view using a single camera and a single mirror, while simultaneously achieving a single effective viewpoint, was fully investigated in [8, 1]. The only such systems correspond to mirror which are surfaces of revolution whose generatrix is a conic (that is symmetric in the axis of revolution). Furthermore, the central camera looking in the mirror must be positioned such that its optical center coincides with a focus point of the mirror¹⁰. The different possibilities for such systems are described in detail in [1] – one can distinguish trivial and degenerate cases from practically useful ones, as follows.

The first trivial/degenerate case was mentioned just above, a planar mirror. Its generatrix is a line orthogonal to the axis of revolution and the camera may be positioned anywhere. The next case corresponds to the generatrix being a pair of lines: the mirror surface is thus a circular cone¹¹. A cone possesses a single focus, its apex. This is thus a degenerate case since a camera located at the cone's apex does actually not see the reflection of the scene in the mirror. Cone-shaped mirrors were still used to build catadioptric system, but with cameras situated away from the apex, leading to a non-central system, see more below.

The remaining three cases correspond to the generatrix being a parabola, ellipse or hyperbola respectively. See figure 4 for illustrations of the explanations given in the following.

The parabolic mirror has two foci – one “inside” the mirror and the other being the point at infinity of the mirror's axis of revolution. There are thus two possibilities to position the camera. Positioning it at the first focus point is rather impractical since it obstructs a large portion of the effective field of view and since the latter is not omnidirectional but rather narrow. It correspond to the “bounding cylinder” of the mirror: the cylinder whose axis is that of the paraboloid and whose lateral extent is the same as that of the mirror. The second possibility is more interesting: the idea of positioning a camera at a viewpoint at infinity imposes that the camera realize an affine projection, i.e. where all viewing rays are parallel to one another and parallel to the mirror's axis. In practice, this can be achieved with an axis-aligned telecentric lens. For the reason explained above, the camera is positioned outside the parabolic mirror, cf. figure 4. The single effective viewpoint of this catadioptric system is the first focus point: indeed, when back-projecting image points from the affine camera and off the mirror, one obtains lines that are all incident with the focus point.

The second regular case consists in using an ellipsoidal mirror. Both foci lie inside the mirror shape. In practice, one would of course use a truncated ellipsoidal mirror surface, otherwise the camera

¹⁰Note that the type of camera is irrelevant as long as it is central, i.e. it may in principle be a perspective camera, a perspective camera with distortions, a central fisheye, etc.

¹¹Only one of the two mathematical cone lobes is used.

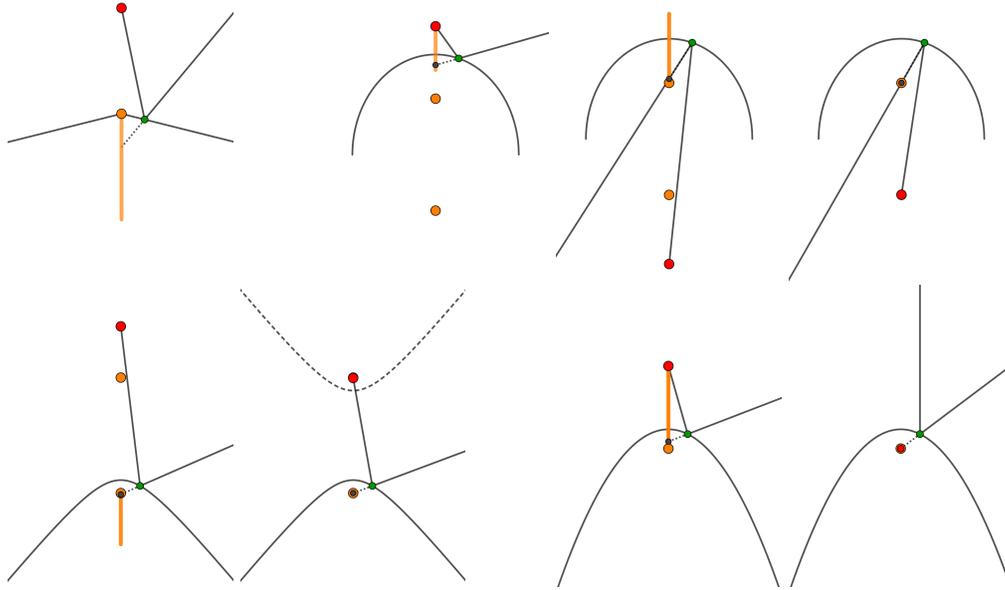


Figure 4: Investigating central and non-central catadioptric cameras. Each figure shows a case where a camera, represented by the optical center in red, is gazing at a mirror. The back-projection along a direction emanating from the optical center and through a point on the mirror (in green), is shown. The orange dots show the focus points of the mirror surface. The main aspect of interest here is where the back-projected rays intersect the mirror axis. In the upper left example (which corresponds to a planar section of a cone-shape mirror), the orange segment shows the locus of all possible such intersection points (when considering all possible back-projection directions). The fact that this locus is a segment and not a single point, means that this catadioptric system is non-central. The three figures on the upper right correspond to an elliptical mirror. Only when the camera is positioned on one of the focus points, does the locus collapse to a single point (identical with the other focus of the ellipse) and the system is a central one (right-most example). The two figures on the lower left show a hyperbolic mirror (the second one shows the two sheets of the mathematical hyperbola, whereas the actual mirror embodies a single one of these). Again, only if the camera is located at a focus point, the system is central. The lower right concerns a parabolic mirror. In the second case, the system is central: the camera is “situated” at the parabola’s focus point that is at infinity (thus not shown) – it must thus carry out an orthographic projection. Again, in this case, the above mentioned locus collapses to a single point – the focus point that is “inside” the mirror.

would not see the scene outside the mirror. Still, this is, like the first parabolic case, of little practical interest, due to a limited useful field of view. Note that the special case of a spherical mirror is actually fully degenerate. Here, the two foci collapse into a single one (the sphere center) and a camera situated there would see nothing else than itself.

The last remaining case is that of a hyperboloidal mirror. To be precise, in practice one uses a mirror corresponding to only one of the two sheets of a hyperboloidal surface. This has two foci, one inside and one outside the mirror’s concave part. Positioning the camera on the second focus point is advantageous since the camera’s reflection occupies a much smaller portion of the final image than in the reverse case.

Overall, there are in general thus two practically useful designs for single-mirror and single-camera central catadioptric systems, based on paraboloid and hyperboloid mirrors. Let us briefly examine the associated geometry of projection and back-projection. Whereas for general camera–mirror system, projection involves (explicitly or implicitly) a search for a point on the mirror surface such that the

scene point's reflection in the tangent plane at that point is directed towards the optical center (cf. section 3), the situation is considerably simplified for central catadioptric cameras. Namely, it suffices to compute the line joining the scene point and the mirror's second focus point, to intersect that line with the mirror and to project that point into the perspective camera located at the other focus (or affine camera in the parabolic case)¹². Back-projection works similarly: one first back-projects the image point to a camera ray, intersects this with the mirror and then computes the line joining the second focus and this intersection point¹³.

While the geometric principle of projection and back-projection is the same for the parabolic and hyperbolic cases (and for the elliptic one too for that matter), it involves a focus point at infinity in the parabolic case and correspondingly, an affine camera instead of the perspective one used in the other cases. Note that a unified model that follows the same operations but that only handles finite entities, the so-called sphere model, has been proposed in [10, 2]. It simplifies the algebraic representation and geometric analysis of central catadioptric cameras.

Let us now briefly study non-central catadioptric systems, in particular systems using mirrors of the same type as above (surfaces of revolution with a conic as generatrix) but where the camera looking at them is not positioned at a focus point. A first observation is that if the camera's optical center lies on the mirror's axis of revolution, then the system is of the axial type: all back-projection rays are incident with a line, the axis of revolution, or rather, in general, a segment of that line. This is due to the following fact: since the optical center lies on the mirror axis, any back-projection ray is coplanar with it, and since the mirror is a surface of revolution, the reflection of that ray in the mirror gives rise to a (half-) line that is again coplanar with the mirror axis. Hence, all final back-projection rays are incident with the mirror axis.

A few examples of this type are studied in figure 4: the figure shows in particular the locus of effective viewpoints on the mirror axis, i.e. the part of the mirror axis that is incident with back-projection rays of the system. In some cases, the locus is a line segment whereas in others, it is a half-line or composed of two half-lines (not shown in the examples since corresponding to impractical settings). Note that in the case of a cone-shaped mirror for instance, an alternative choice to represent the viewpoint locus is possible instead of a line segment on the mirror axis: indeed, all back-projection are also incident with a particular circle centered in the mirror axis and "located" outside the cone [1].

If the optical center of the camera is not located on the mirror's axis, then the system is in general no longer axial and can be considered as fully non-central. Note that there are different "degrees of non-centrality": the most general one corresponds to the so-called oblique cameras [30] where no two back-projection rays intersect one another.

Let us close this section by referring to section 1 where properties of central versus non-central cameras were already discussed. Further differences are explained throughout the following sections, concerning structure from motion problems, epipolar geometry and dense stereo matching.

5 "Outer" Geometry: calibrated cameras

In this chapter we make the assumption that camera rays are straight (half-)lines, i.e. light rays that form the camera image travel on straight paths before entering the aperture. More general situations

¹²To be precise, there are two intersection points and thus two mathematical solutions for the projection.

¹³Again, two mathematical solutions exist in general. In the parabolic case, one of the two solutions is always the same since the optical center – the point at infinity of the mirror axis – lies on the parabola. Hence, the mirror axis is always one of the two mathematical solutions to back-projection.

occur for instance in airborne or satellite imagery, where atmospheric refractions may require to loosen this assumption (an issue studied already decades ago in photogrammetry). Similar other examples concern scenes which are composed of multiple different media, such as cameras looking from the air into water, possibly through a curved glass window. The subjects handled in this section for straight camera rays have also been studied for such situations in the literature, giving rise to quite interesting geometric questions and findings [23, 5, 42].

Given the scope of this chapter, we now return to the case of straight camera rays. We ask the question how to represent and infer information on the relative positioning between cameras and/or entities in the scene. This question has a very long history, which can be traced back at least to the Italian Renaissance and, for mathematical treatments, by at least about 300 years. In the following we give a very concise overview of what we consider to be some main versions of this question.

Given an image of a scene and a particular point in that image, where could the original point in the scene possibly be located? Since in this section we assume a fully calibrated camera, the immediate answer is trivial and it corresponds to the issue of back-projection already introduced: the scene point must be located somewhere on the (half)-line that can be computed/constructed from the calibration information. This is all one can say in the absence of further information. Additional information which may allow to pin down the scene point’s location more precisely, is for example exploited in “bad weather imaging”, where the amount of blur present around the image point, caused by fog in the scene, may give clues as to the depth of the scene point [26]¹⁴. In this chapter we do not consider such cases further and only use the “pure” geometrical information conveyed by (mathematical) points.

Is it possible to precisely locate an object in 3D from a single image and if yes, what information is required to do so and how to solve this problem mathematically? Here, an object is assimilated to a set of points. The main variant of this problem is nothing else than pose estimation or, as it is called in photogrammetry, resection: the required information is knowledge about the object’s shape (i.e. the relative position of the object points). The minimal case where a general solution is possible is an object consisting of three points only. The “shape” of such an object is for example fully represented by the three pairwise distances between the points. The first work known to us that reports this problem and sketches a mathematical solution is by Lagrange, in 1773 [22]. He already showed that the problem can be reduced to finding the roots of a quartic polynomial and also sketches an iterative numerical procedure. While Lagrange did not explain in detail the construction of this quartic polynomial, he very likely had the complete solution. In 1841, a complete analytical solution was eventually provided by Grunert [13].

How can we describe the pose estimation problem geometrically? One way of seeing it is as follows. First, we back-project the three image points, giving three lines in 3D. Second, we try to displace the object (a triangle) in 3D, by translating and rotating it, such as to put it into a position where each of its three points comes to lie on the associated camera ray. This is something one can really try at home: arrange 3 spaghettis any way you like and try to find out how to place some triangular object in the described way.

In general, there are up to 8 different solutions to this problem. In the case of a central camera, as considered by Lagrange and Grunert, there are 4 pairs of mirror solutions (see figure 5). Whereas for

¹⁴Such approaches require of course further assumptions, such as on the homogeneity of the fog and the texture of the scene. Also, more generally, the recent years have seen an increasing number of works where different cues are exploited for single-image 3D modelling through machine learning.

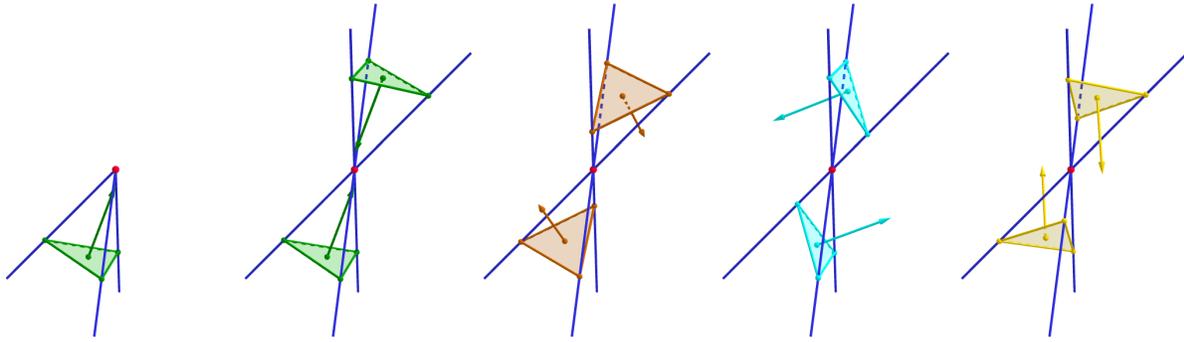


Figure 5: Illustration of the 3-point pose problem. First: a central camera (only the optical center is shown, in red) gazing at 3 points. The camera rays are shown in blue. The 3-point pose problem is to estimate the pose (position) of the 3 points, given only the camera rays (or, equivalently, the image points and the camera calibration) and knowledge of the pairwise distances between the original 3D points. Second: one solution is of course the original pose – a second mathematical solution is given by the mirrored pose (3D points symmetrically “behind” the camera). The remaining figures show three other pairs of solutions. Without further information (such as on the inclination of the normal vector of the plane spanned by the points), one cannot in general disambiguate the solutions.

non-central cameras, where the camera rays do not intersect in some common point, 8 entirely different solutions may exist and where, incidentally, the analytical formulation comes down to finding the roots of an 8th degree polynomial [6, 29, 33].

Let us make a few complementary remarks on the pose estimation problem. First, it can be stressed that the sketched minimal solutions are applicable to any type of central respectively non-central camera. For the central case for instance, it does not matter if the back-projection rays stem from a perspective camera, a fisheye, a catadioptric, or any other central camera. Second, as is the case with many other minimal problems, the number of admissible solutions can in general be reduced when more than the minimum number of data (here, points) are available. More in this issue is discussed in the last part of this section. Finally, various special cases of the pose estimation problem have been covered in the literature, for example the cases of a linear object or of an object consisting of more than three coplanar points, the case where the unknown pose has fewer degrees of freedom than six, etc.

Is it possible to estimate the motion of a camera just by taking images of an unknown scene?

This important problem is called relative orientation in photogrammetry and motion or ego-motion estimation in computer vision. Maybe the first general solution is due to Kruppa who, in 1913, showed that five point correspondences between two images of a static scene, are sufficient to solve this problem [21]. The minimal 5-point problem is algebraically rather complex and only in 2004 the first approach was published which provided only the up to 10 theoretical solutions (previous approaches gave a higher number of solutions, some of which being spurious) [28]. Note that these approaches are designed for calibrated central cameras; for non-central ones, the minimal case is six point correspondences [37].

Geometrically, the motion estimation problem can be sketched as follows. Given the calibration information, the image points can first be back-projected. This gives two sets of 3D camera rays, one per image. It then suffices to align these two sets by rotating and translating them one against the other such that camera rays that belong to corresponding image points, intersect one another.

Unlike pose estimation, motion estimation is subject to a fundamental difference between central and

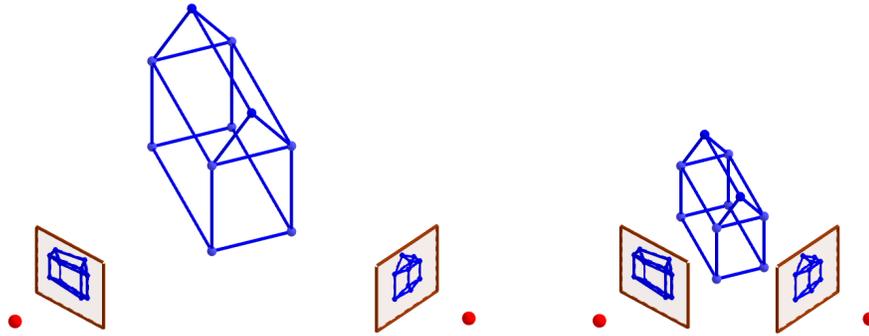


Figure 6: Scale ambiguity of relative motion estimation with two central cameras: the relative motion can be estimated at best up to the scale of the translation between the cameras. The 3D reconstruction of the scene that can be computed using the estimated motion, is affected by the same scale ambiguity. See an animation in the video `Scale_Ambiguity.mp4`.

non-central cameras. Let us first note that the problem has in principle six unknowns (three each for rotation and translation). However, if both cameras are central, then the translation can only be estimated up to scale, i.e. the direction can be determined but not the extent of the translation as can be seen from figure 6, if we have one possible relative motion, then moving the cameras along the baseline between the optical centers will leave the fact unchanged that corresponding camera rays intersect. This is the reason that five correspondences is the minimum case here. As for non-central cameras, this so-called scale ambiguity vanishes in general. This implies on the one hand that at least six correspondences are required, on the other hand that camera motion is not only estimated up to scale, but precisely. This potentially strong advantage of non-central cameras has to be weighed with care though: if the camera is only “slightly non-central”, that is if the camera rays pass all through some small volume in space, then the accuracy of the estimated scale of translation might be poor. More on this issue is explained in section 3.6 of [38].

Note that, like for pose estimation, various special cases have been studied in the literature. The special case of a planar object was already solved by Schröter in 1880 (paragraph 45 of [35]). In [28], Nistér provides a method that directly solves the relative motion between three images, which gives in many cases stabler results, and especially, works for planar scenes, in which case the two-view method fails.

Let us finally remark that motion estimation is closely linked to epipolar geometry (see section 7). Especially for calibrated cameras, knowing the epipolar geometry is equivalent to knowing the relative motion between two images.

Triangulation – Reconstructing 3D points. The last problem discussed here is that of reconstructing 3D points from image correspondences between calibrated cameras whose relative motion is known or has been estimated (to give for instance a 3D model such as the one shown in figure 6). The minimal case corresponds obviously to two images. The first optimal method for this seemingly simple problem has been provided in [14], along with suboptimal methods for using correspondences in multiple images.

Some remarks. For the construction of complete and detailed 3D models, one usually requires several to many input images. Many 3D modeling pipelines have been proposed in the literature. They are usually based on building blocks such as methods for the above or similar problems, complemented by approaches for extending them to multiple images.

In the remainder of this section, we make a few remarks on the numerical methods used to solving these problems.

Above we explained the minimum amount of information required to solve the examined problems. In practice, one usually has more data at one's disposal, which opens the door to different ways of estimating the unknowns of the problem. We may distinguish three main types of approach.

First, the so-called minimal methods, working exactly with the minimally required amount of information (three points for pose estimation, five for motion estimation, two views for triangulation, and likewise for the mentioned special cases). These are usually expressed through (systems of) polynomial equations, see for instance [28, 14, 13]. Investigating minimal cases is of at least two-fold interest. On the one hand, this provides theoretical insights into studied problems, their complexity, the number of theoretically possible solutions, degenerate configurations where the problem is unsolvable, etc. On the other hand, minimal solvers are very relevant in practice in case the input may contain outliers, that is wrong correspondences between points. As is well known, such outliers, if not taken care of, may lead to completely erroneous numerical solutions to an estimation problem. One routine way of handling the outlier issue is to have recourse to random sampling approaches. The principle of the classical RANSAC (RANdom SAMple Consensus [9]) approach and its many variants is to repeatedly draw minimal random samples from the input data (e.g. triplets of points for pose estimation), to estimate hypothetical solutions to the problem from them and to check these against the remaining input data. With sufficiently many random samples the probability of obtaining at least one sample with only inlier data is high – and in that case the consistency check with the other input data will signal a good consensus since, roughly speaking, most inliers will “signal agreement” with the hypothetical solution provided by that sample. The fact that a sample may provide multiple hypothetical solutions (such as up to four in the case of pose estimation), is handled trivially by checking the consistency of each of these solutions with the other input data. RANSAC is just one approach for robust estimation; for other possibilities, see for instance [19, 43].

A second type of approach consists in trying to estimate a solution by using more than the minimum required data at once through solving a system of linear equations. For instance, how to approach pose estimation in this manner was explained in [32] and as for the triangulation problem with an arbitrary number of input images, see [14]. Such approaches have been quite popular in computer vision as they are computationally rather simple. A drawback however is that for most problems, the linear solution is suboptimal, as explained below.

The third main type of approach aims at a rigorous handling of uncertainty in the data. Ideally, one tries to estimate the uncertainty of all input data and propagate it throughout the estimation process. For instance, when extracting interest points in images, one may estimate, besides the point positions, covariance matrices for the latter. The computation of problem variables (camera pose, relative motion, etc.) can then be formulated as a maximum a posteriori (MAP) estimation problem, where the likelihood is evaluated against these covariance matrices on the input data (or any other representation of their uncertainty). For most geometric computer vision problems, this then comes down to solving a non-linear optimization problem. The general problem where the location of 3D scene points as well as camera poses are to be estimated, is traditionally called bundle adjustment, see [43] for a general overview. As mentioned, the objective function for these MAP formulations is non-linear for most problems. Conversely, the linear formulations mentioned above, are equivalent to solving maximum likelihood problems whose objective functions do not assess the quality of the estimates against the uncertainty in the input data – which is one explanation for their suboptimal performance.

Besides the above three types of approach, let us mention that for some problems, global optimization methods have been devised, see for instance [15].

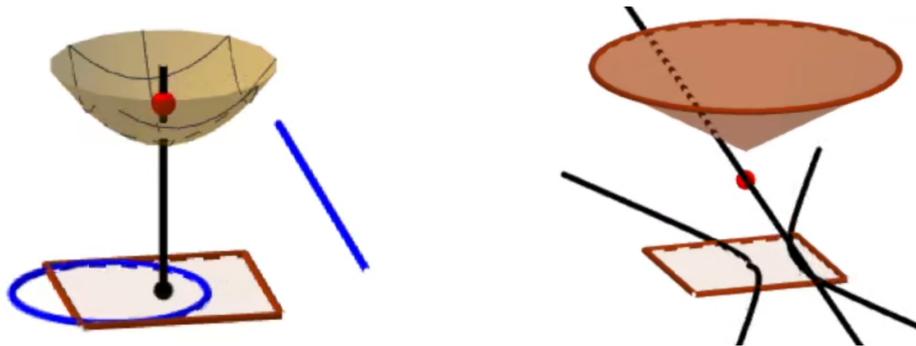


Figure 7: Two examples of a catadioptric line image. Left: with a central catadioptric camera, line images are circles. Right: this system based on a cone-shaped mirror is non-central – the line image is not a conic as with all central catadioptric systems, but a higher degree curve (note for instance the small cusp in the curve on the left).

6 “Inner” Geometry: Images of Lines

One of the most interesting and important aspects of the geometry of omnidirectional cameras, is how lines appear in the image. Whereas the image of a point is obviously a point itself under the assumptions stated earlier (or a set of points), the omnidirectional image of a line is in general not a line (see figure 7).

The study of line images is interesting in several respects. First, they tell much about the calibration of an omnidirectional camera and using lines is an attractive approach for calibration. Second, the geometry of line images is directly related to epipolar geometry and thus to (dense) binocular stereo-vision. In the following, we explain the link between line images and camera calibration; the issue of epipolar geometry is handled in the next section.

Let us consider the example of central catadioptric cameras. It is well known that for these, the image of a line is a conic [3]. To be precise, this is only true in general if the camera looking at the mirror is a perspective one: otherwise, for instance if the camera is subject to radial or other distortions, the line image is no longer a conic, but a conic “distorted” accordingly. But let us adopt the assumption of a perspective camera and thus the fact that line images are conics.

Let us provide a simple counting argument to explain why line images contain information on the camera calibration. It is well known that the set of lines in 3D has four degrees of freedom. Now, consider a central camera and the image of some 3D line L . It is clear that any other line that lies in the plane spanned by L and the camera’s optical center, gives rise to the same line image as L (with the exception of lines going through the optical center, which constitute a singular case). Such lines represent a set with two degrees of freedom. This suggests that the set of images of all 3D lines in a given camera, only has two degrees of freedom ($4 - 2$). If the camera is perspective (and looking directly at the scene, i.e. not through a mirror or other device), then line images are lines themselves. The set of lines in the image plane has two degrees of freedom. This illustrates that for perspective cameras, images of lines do not convey any information on the camera’s calibration, since no excess information is contained in them. Or in other words: for any line in the image plane, there exist lines in 3D that are projected onto that image line. Hence, knowing lines images is not providing information useful for calibrating perspective cameras.

The situation is different for central catadioptric cameras. As already mentioned, line images are conics. The set of all conics has five degrees of freedom but as already explained, the set of line images only has two of them. In other words, not every conic in the image plane is a possible line

image. Hence, knowing line images does tell something about the camera's calibration. Calibration methods for catadioptric cameras that exploit this fact, have been proposed for instance in [11, 3].

The situation is analogous for other types of omnidirectional cameras and for non-perspective cameras in general. For instance, using lines to calibrate non-perspective distortions in regular cameras, is a classical approach called plumline calibration [4]. As for perspective cameras with radial or other distortions, plumline calibration allows in general to estimate all calibration information about the camera, up to the perspective part (i.e. focal length, principal point, aspect ratio and coordinate axis shearing). As for omnidirectional cameras, line images may allow to recover the entire calibration information. A formal study of the conditions in which full calibration is possible from line images does not exist to the knowledge of the author, but would be an interesting endeavor.

So far, we have only discussed the case of central (catadioptric) cameras. For non-central ones the situation differs, in that the images of any two 3D lines may differ from one another, hence that the set of line images now may have up to four degrees of freedom and not only two. Line images may still convey information useful for calibration. A general study of plumline calibration for non-central cameras is another topic worthy of exploration.

7 Epipolar Geometry

Epipolar geometry is a fundamental topic in computer vision and photogrammetry¹⁵. It essentially arises from the following question: given two images of a scene and a point in one of these, where could the corresponding point possibly be located in the other image? This question is one ingredient of the image matching problem, which is a pre-requisite to most structure-from-motion tasks and dense stereovision. Studies of epipolar geometry usually concern at least two aspects of the question. First, what can one say about the nature of the locus of corresponding points? For perspective cameras for example, the locus is a straight line whereas for central catadioptric cameras, it is a conic, see more below. Second, what is needed to compute that locus? How to compute it from information on the calibration and the relative pose of the two cameras? Or how to compute it from a (potentially small) set of already available image matches? In the following, we only address the first aspect, followed by a discussion of implications for dense binocular stereo and rectification.

7.1 Nature of epipolar geometry

We address here the first question: given a point in one image, what can one say about the nature of the locus of corresponding points in the other image? The immediate answer directly follows from the considerations on line images exposed in the previous section and on back-projection, see section 3. Namely, we can simply decompose the problem into two parts. First, we ask the question where the original 3D point in the scene could possibly be located – this is simply the back-projection problem. As already seen, this gives rise to a (half-) line in 3D: without any additional information, the original point may be located anywhere on it. Second, the locus of corresponding points in the second image is then nothing else than the image of that (half-) line.

Let us call that line image the *epipolar curve* of the first image point, in analogy to *epipolar lines* of perspective images. The first answer to our question is thus: the corresponding point in the other image must lie on an epipolar curve and the nature of that curve is nothing else than the nature of

¹⁵The oldest work known to us, is [40] – the actual term epipolar geometry was introduced much later.

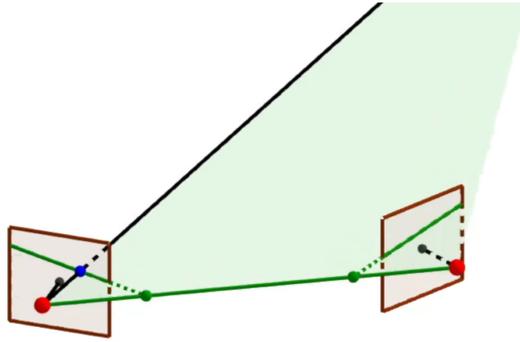


Figure 8: Epipolar geometry of two perspective cameras. The image point in blue gives rise to a back-projection ray, whose image in the second camera is an epipolar line. The epipolar plane is the plane spanned by the back-projection ray and the optical center of the second camera – the epipolar line is the intersection of that plane with the image plane. The epipole of each camera (green point) is where the baseline between the cameras intersects the image plane (here, the epipoles are situated outside the image area, but this is irrelevant for most considerations). All possible epipolar lines are incident with the epipole, i.e. they form a line pencil, with a single degree of freedom.

line images of the second camera, i.e. conics for central catadioptric cameras, lines for perspective cameras (cf. figure 8), etc.

It is important to realize that the nature of epipolar curves in one image depends only on the characteristics of the camera acquiring that image and not on the those of the other camera. In other words, whatever type the other camera is does not influence the nature of these epipolar curves, even if the other camera is non-central or otherwise different from the considered one.

The discussion so far concerns the nature of individual epipolar curves. To fully study epipolar geometry, one should also examine the characteristics of the set of all possible epipolar curves. Here, the type of the other camera does play its role. . . Let us explain this: whichever type the other camera is, back-projection of individual image points always gives rise to lines in 3D, hence the observations made above on the nature of epipolar curves. However, the composition of the set of all possible back-projection lines obviously depends on the type of camera. For a central camera, all back-projection lines go through the optical center whereas for non-central cameras, this is by definition not the case.

These observations lead to several special cases depending on the types of the two cameras. A few of them are examined in the following. First, the classical case of two perspective cameras (see the accompanying video `Epipolar_Geometry_Standard_Rectification.mp4`). Let us look at the set of back-projection lines of the first camera: this is a (subset) of a bundle of lines, all going through the optical center. The bundle of lines has two degrees of freedom. However, the set of the images of these lines in the other camera only has one degree of freedom. To see this, let us consider any one line in the bundle and the plane spanned by that line and the optical center of the second camera (a so-called epipolar plane). That plane contains a one-dimensional subset of lines in the bundle. All of them get projected to the same line image in the second camera. Overall, the lines in the bundle may be “partitioned” in a one-dimensional family of sets of lines lying in one epipolar plane each and with each set having the same line image. Thus the set of epipolar lines in the second camera is one-dimensional only. Furthermore, as is well known, all epipolar lines go through one particular point, the epipole, which is nothing else than the perspective projection of the first optical center in the second camera.

What if we replace the first camera by any non-perspective but still central camera? Well, as far as this discussion is concerned, nothing changes: epipolar curves in the second camera are still the same

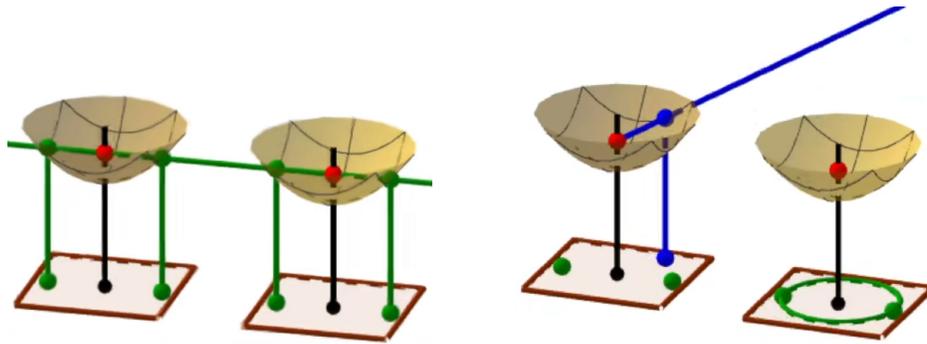


Figure 9: Epipolar geometry of two para-catadioptric cameras. Left: the two epipoles in each image are the two (mathematical) image points of the other camera’s effective optical center (in red). Right: Epipolar curve (circle in this case) in the second camera, being the image of the back-projection ray associated with the shown image point in the first camera. More on this is shown in the video `Epipolar_Geometry_Para.mp4`.

set of epipolar lines, all incident with the same epipole.

How about the epipolar geometry if the second camera is for example a central catadioptric one and the first one any central camera? Then, as already hinted on, the epipolar curves in the second camera form a two-degree-of-freedom set of conics. What about the epipole? In this case, since central catadioptric cameras map each 3D point to two image points, there are a pair of epipoles (cf. figure 9) and all epipolar conics are incident with both of them. As is clear by now, this observation is independent of the precise type of the first camera, as long as it is central.

Similar findings hold for other types of central cameras, e.g. fisheyes: the nature of the epipolar curves is that of line images and the set of epipolar curves are a particular subset of admissible line images, all being incident with the epipole (or epipoles in case the camera maps 3D points to more than one image point each).

Let us now consider the case where the second camera is perspective but the first one non-central. A first example is that of a catadioptric camera with a cone-shaped mirror such that the camera looking at the mirror is situated on the mirror’s axis. As shown above, this is a non-central device. All back-projection lines are incident with the mirror’s axis and more specifically, with the line segment representing the caustic of the system (cf. figure 4). Now, if the perspective camera is also positioned on the mirror’s axis, then the back-projection lines are projected to a one-dimensional set of image lines, all incident with the point that corresponds to the “image” of the mirror axis, a point that can be considered as epipole. The situation differs as soon as the perspective camera moves away from the mirror’s axis. Now, the back-projection lines are mapped to a two-dimensional subset of the lines in the perspective image plane. To be precise, the epipolar lines are all lines in the perspective image plane which are incident with the line segment that is the image of the catadioptric system’s caustic. There thus no longer exists a distinct epipole; at best, one may consider the said line segment as playing a somewhat analogous though weaker role as that of a “regular” epipole.

Figure 10 shows examples of the epipolar geometry of two non-central cameras, catadioptric cameras with a cone-shaped mirror¹⁶. In the first case, when the two cameras are arranged on top of each other, the epipolar geometry is similar to that of perspective cameras, i.e. epipolar curves are straight lines and there exist epipoles. This is no longer true in the second case.

¹⁶See also videos `Cones_Aligned_Epipolar_Geometry.mp4` and `Cone_Projection_And_Noncentral_Epipolar_Geometry.mp4`.

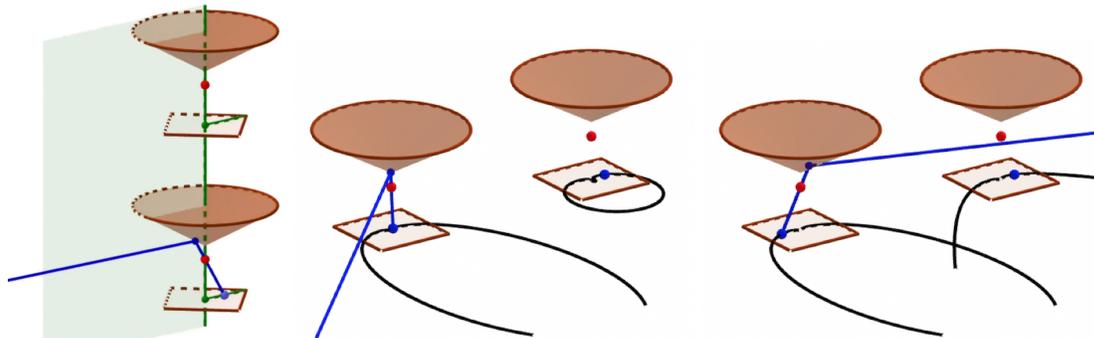


Figure 10: Epipolar geometry of two non-central cameras – here, catadioptric systems with cone-shaped mirrors. Left: if the two are placed on top of each other, the epipolar geometry is similar to that of perspective cameras. For ease of comprehension, only one half of the epipolar plane is shown. Middle and right: Now the two systems are placed side by side. The figures show the epipolar curve in the first camera, associated with the image point in the second camera. They then consider a point in the first image, located on this epipolar curve, in two different positions. The epipolar curves in the second camera, associated with these two image points, differ from one another. This indicates that epipolar curves are no longer in a one-to-one correspondence here, unlike with central cameras.

An interesting question is if such a “weak” epipolar geometry is still useful for image matching. There are two answers to this. First, if the task at hand is to find the corresponding point of one individual image point in the catadioptric image, then the answer is positive: the corresponding point must lie on the epipolar curve associated with the point in the first, catadioptric, image, and it does not matter that this epipolar curve is part of a weak epipolar geometry in the sense described. If however the task to be carried out is dense stereo matching, then the answer turns out to be negative in general. To examine this, we will first return to the case of two perspective images and recap how dense stereo matching can be, and usually is, carried out in that case.

7.2 Dense stereo matching and rectification

The standard pipeline for dense stereo matching is to first rectify the image pair and then to perform so-called scanline matching. Let us describe the rationale behind and workings of this procedure. The first essential aspect is that epipolar lines in two perspective images come in pairs: consider some point in the first image and the associated epipolar line in the second image. It turns out that for any point on that line, the associated epipolar line in the first image is always the same line (the line spanned by the first point and the epipole in the first image). There is thus a one-to-one correspondence: to each epipolar line in one image corresponds exactly one epipolar line in the other image. So, to find the matches to points on one epipolar line, it suffices to search among the points on the associated epipolar line in the other image. This means that the “complexity” of dense stereo matching is lower than it might appear at first: instead of doing a one-dimensional search each for a two-dimensional set of image points, it is possible to partition the problem into a one-dimensional set of problems, each of which consists in matching all the points on corresponding epipolar lines to one another in a single process.

The main advantage of the geometry at hand comes into play however through the way one can implement this last problem efficiently – this is not so much of a geometric issue but more of a computational one, in that it relies essentially on organizing computations and memory access efficiently. Let us first consider the initial situation where the goal is to find the matching point to a point in the first image and where the epipolar line is not axis-parallel. Point matching in general relies on

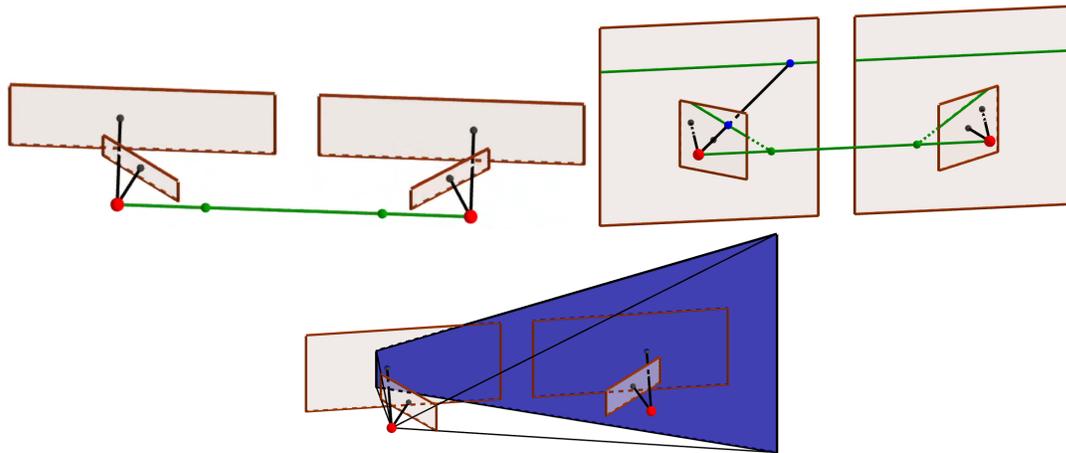


Figure 11: Classical stereo rectification for perspective images. Upper left: one defines two coplanar virtual image planes that are parallel to the baseline. Upper right: one “projects” epipolar lines onto these virtual image planes. There, they will be axis-aligned and occupy identical image rows, allowing for efficient implementation of dense binocular stereo. Bottom: The required area for the first rectified image (camera slightly more panned inside than in the other figures). See also the video `Epipolar_Geometry_Standard_Rectification.mp4`.

comparing greylevels in windows around image points and computing some similarity measure from them. To find the matching point along the epipolar line, the basic procedure is thus to sample the latter by considering successive points on it, computing the similarity measure for each of those, and then choosing the point that maximizes the similarity measure.

A first main source of the computational cost of this is memory access: each time a new point on the epipolar line is considered, one must access the greylevels of the pixels around it and this requires costly non-local memory accesses. A solution to circumvent this is to carry out a pre-processing where greylevels of each image are re-arranged in memory in the following way: they are stored in a matrix such that greylevels of points on the same epipolar line appear in the same row. Further, neighborhood relations should be preserved, meaning that greylevels of neighboring points (within an epipolar line across neighboring lines) appear next to each other in that matrix. With this way of organizing the image information, the cost of memory access for the actual matching stage is much reduced: successively exploring all points on an epipolar line can be done through local memory accesses (simply put, by exploring sequentially arranged data in memory)¹⁷.

This description of how to rearrange the information contained in the images for efficient stereo matching is actually just an alternative way of explaining the rationale of the classical process of stereo rectification, illustrated in figure 11.

A significant drawback of this method for rectification¹⁸ is that they cannot be applied when any of the epipoles lies inside the image area and that they become impractical when epipoles get close to the image area. The reason is that the required image area for the rectified images, becomes prohibitively if not infinitely large (cf. figure 11). This situation is of course frequent with omnidirectional images – with fields of view extending 180° and when epipoles exist (e.g. with central cameras), at least one of them will always lie within the image area. . .

¹⁷A second important way of reducing computational cost of dense stereo matching is through avoiding redundant computations. For instance, when computing the similarity measure for a pair of points and for the pair consisting of the two points’ right neighbors, some if not most of the computations done in the two cases will be identical. Through appropriate bookkeeping the overall computational cost can be easily decreased dramatically.

¹⁸and more general ones, such as reasoning on homographies that map epipoles to points at infinity in the image [16]

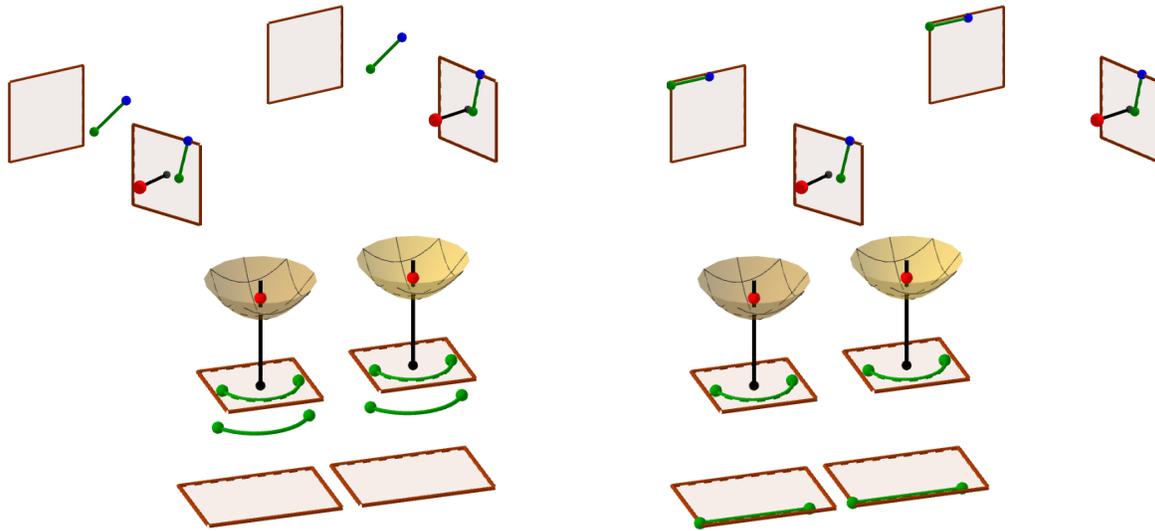


Figure 12: Generic stereo rectification. Upper left: Epipolar segments are “cut out” from original images and “pasted” into the rectified ones. Upper right: After the pasting. Bottom: the same, for paracatadioptric images. See videos `Epipolar_Geometry_General_Rectification.mp4` and `Epipolar_Geometry_Para.mp4` for animations and more explanations.

Alternative methods for rectification circumvent this problem, in an altogether simple manner [31, 12]. Their rationale is essentially as follows. Instead of rectifying epipolar lines through a perspective projection onto a virtual image plane, one simply “cuts out” the line segments occupied by epipolar lines in the image area, and “pastes” them on rows of the rectified images to be generated, as illustrated in figure 12. All that needs to be done is to paste corresponding segments in the two images to corresponding rows in the rectified images, and to take care that pixels that neighbor each other in the original images, end up as neighbors in the rectified ones.

This idea, first developed for perspective images, can be readily extended to omnidirectional ones. The only essential additional step is that subsets of epipolar curves instead of line segments are concerned; it suffices to “unroll” these in order to paste them into the rectified images, as illustrated in figure 12. With this simple procedure, one thus generated rectified image pairs that can in principle be processed using standard dense binocular stereo methods. Naturally, when exploiting the stereo matching results in order to generate, for instance, a dense depth map, one may need to map the matches obtained for rectified images, back to the original images, so that calibration information for these can be used for 3D reconstruction.

Let us close this section by reminding that the application of this rectification principle requires that epipolar curves are in one-to-one correspondence, in the sense explained above. This is always the case when the cameras are central and sometimes even when they are non-central, such as with cone-based catadioptric cameras on top of each other, as illustrated above. With non-central cameras in general position however, this property breaks down. While epipolar curves continue to exist, they are in general no longer in one-to-one correspondence across the images. A general study on all possible stereo images, is provided in [36].

8 Closing remarks

This chapter has revisited some fundamental geometrical aspects of camera models, with an emphasis on omnidirectional ones. Many topics were not discussed, such as multi-view matching relations (for

more than two images), calibration or self-calibration. These, as well as algebraic expressions for the concepts presented, can be readily found elsewhere, for instance in several chapters in this book.

References

- [1] S. Baker and S.K. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):1–22, 1999.
- [2] J.P. Barreto and H. Araújo. Geometric properties of central catadioptric line images. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 4, pages 237–251, 2002.
- [3] J.P. Barreto and H. Araújo. Geometric properties of central catadioptric line images and their application in calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1327–1333, 2005.
- [4] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.
- [5] V. Chari and P. Sturm. Multi-view geometry of the refractive plane. In *British Machine Vision Conference, London, UK*, 2009.
- [6] C.-S. Chen and W.-Y. Chang. On pose recovery for generalized visual sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):848–861, 2004.
- [7] S.E. Chen. Quicktime VR – an image-based approach to virtual environment navigation. In *Proceedings of SIGGRAPH, Los Angeles, USA*, pages 29–38, 1995.
- [8] D. Drucker and P. Locke. A natural classification of curves and surfaces with reflection properties. *Mathematics Magazine*, 69(4):249–256, 1996.
- [9] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381–395, 1981.
- [10] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3):223–243, 2001.
- [11] C. Geyer and K. Daniilidis. Paracatadioptric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):687–695, 2002.
- [12] C. Geyer and K. Daniilidis. Conformal rectification of an omnidirectional stereo pair. In *Proceedings of the Workshop on Omnidirectional Vision and Camera Networks, Madison, USA*, 2003.
- [13] J.A. Grunert. Das pothenot’sche Problem in erweiterter Gestalt; nebst Bemerkungen über seine Anwendung in der Geodäsie. *Archiv der Mathematik und Physik*, pages 238–248, 1841.
- [14] R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

- [15] R.I. Hartley and F. Kahl. Global optimization through rotation space search. *International Journal of Computer Vision*, 82:64–79, 2009.
- [16] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [17] R. Hicks. Designing a mirror to realize a given projection. *Journal of the Optical Society of America A*, 22(2):323–330, 2005.
- [18] R.A. Hicks and R.K. Perline. Equi-areal catadioptric sensors. In *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pages 13–18, 2002.
- [19] P.J. Huber. *Robust Statistics*. John Wiley & Sons, Ltd., 1981.
- [20] U. Iwerks. Panoramic motion picture camera arrangement, U.S. patent 3,118,340, 1964.
- [21] E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. In *Sitzungsberichte der mathematisch-naturwissenschaftlichen Klasse der kaiserlichen Akademie der Wissenschaften, Abteilung II a*, volume 122, pages 1939–1948, 1913.
- [22] J.-L. Lagrange. Solutions analytiques de quelques problèmes sur les pyramides triangulaires. *Nouveaux Mémoires de l’Académie royale des Sciences et Belles-Lettres*, pages 149–176, 1773. reprinted in 1869 in the 3rd volume of the “Œuvres de Lagrange” edited by J.-A. Serret and published by Gauthier-Villars, pages 661–692.
- [23] H.-G. Maas. New developments in multimedia photogrammetry. In A. Grün and H. Kahmen, editors, *Optical 3-D Measurement Techniques III*. Wichmann Verlag, Karlsruhe, 1995.
- [24] V.S. Nalwa. A true omnidirectional viewer. Technical Report Bell Laboratories Technical Memorandum, BL0115500-960115-01, AT&T Bell Laboratories, 1996.
- [25] S. Nayar and V. Peri. Folded catadioptric cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, USA*, pages 217–223, 1999.
- [26] S.K. Nayar and S.G. Narasimhan. Vision in bad weather. In *Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, Greece*, pages 820–827, 1999.
- [27] R.C. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58(4):261–273, 1988.
- [28] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004.
- [29] D. Nistér. A minimal solution to the generalized 3-point pose problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, USA*, pages 560–567, 2004.
- [30] T. Pajdla. Stereo with oblique cameras. *International Journal of Computer Vision*, 47(1-3):161–170, 2002.
- [31] M. Pollefeys, R. Koch, and L. Van Gool. A simple and efficient rectification method for general motion. In *Proceedings of the IEEE International Conference on Computer Vision, Kerkyra, Greece*, pages 496–501, 1999.

- [32] L. Quan and Z.D. Lan. Linear n-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):774–780, 1999.
- [33] S. Ramalingam, S. Lodha, and P. Sturm. A generic structure-from-motion algorithm for cross-camera scenarios. In *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, pages 175–186, 2004.
- [34] S. Ramalingam, P. Sturm, and S. Lodha. Theory and calibration algorithms for axial cameras. In *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, volume I, pages 704–713, 2006.
- [35] H. Schröter. *Theorie der Oberflächen zweiter Ordnung und der Raumkurven dritter Ordnung als Erzeugnisse projektivischer Gebilde – Nach Jacob Steiner’s Prinzipien auf synthetischem Wege abgeleitet*. B.G. Teubner, Leipzig, 1880.
- [36] S. Seitz and J. Kim. The space of all stereo images. *International Journal of Computer Vision*, 48(1):21–38, 2002.
- [37] H. Stewénius, D. Nistér, M. Oskarsson, and K. Åström. Solutions to minimal generalized relative pose problems. In *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, 2005.
- [38] P. Sturm, S. Ramalingam, J.-P. Tardif, S. Gasparini, and J. Barreto. Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1-2):1–183, 2011.
- [39] R. Swaminathan, M. Grossberg, and S. Nayar. Designing mirrors for catadioptric systems that minimize image errors. In *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, 2004.
- [40] A. Terrero. Topofotografía, ó sea aplicaciones de la fotografía al levantamiento de los planos topográficos. *La Asamblea del Ejército y de la Armada*, V/2/3:31–46, 1862.
- [41] A. Torralba and W.T. Freeman. Accidental pinhole and pinspeck cameras: revealing the scene outside the picture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA*, pages 374–381, 2012.
- [42] T. Treibitz, Y.Y. Schechner, C. Kunz, and H. Singh. Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):51–65, 2012.
- [43] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, Corfu, Greece*, pages 298–372, 1999.