

# A Comprehensive and Versatile Camera Model for Cameras with Tilt Lenses

Carsten Steger<sup>1</sup> 

Received: 14 March 2016 / Accepted: 30 September 2016 / Published online: 22 October 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** We propose camera models for cameras that are equipped with lenses that can be tilted in an arbitrary direction (often called Scheimpflug optics). The proposed models are comprehensive: they can handle all tilt lens types that are in common use for machine vision and consumer cameras and correctly describe the imaging geometry of lenses for which the ray angles in object and image space differ, which is true for many lenses. Furthermore, they are versatile since they can also be used to describe the rectification geometry of a stereo image pair in which one camera is perspective and the other camera is telecentric. We also examine the degeneracies of the models and propose methods to handle the degeneracies. Furthermore, we examine the relation of the proposed camera models to different classes of projective camera matrices and show that all classes of projective cameras can be interpreted as cameras with tilt lenses in a natural manner. In addition, we propose an algorithm that can calibrate an arbitrary combination of perspective and telecentric cameras (no matter whether they are tilted or untilted). The calibration algorithm uses a planar calibration object with circular control points. It is well known that circular control points may lead to biased calibration results. We propose two efficient algorithms to remove the bias and thus obtain accurate calibration results. Finally, we perform an extensive evaluation of the proposed camera models and calibration algorithms that establishes the validity and accuracy of the proposed models.

**Keywords** Camera models · Tilt lenses · Scheimpflug optics · Camera model degeneracies · Camera calibration · Bias removal · Stereo rectification

## 1 Introduction

One problem that often occurs when working on machine vision applications that require large magnifications is that the depth of field becomes progressively smaller as the magnification increases. Since for regular lenses the depth of field is parallel to the image plane, problems frequently occur if objects that are not parallel to the image plane must be imaged in focus. Section 2 describes some of the applications where this problem occurs.

It is well known that the depth of field can be centered around a world plane that is tilted with respect to the image plane. This can be achieved by tilting the lens with respect to the image plane. This is the well-known Scheimpflug principle, which is described in detail in Sect. 3.

When working with tilt lenses, the fact that the ray angles in the object and image space of a lens may differ becomes crucial. Section 4 describes the principles of geometric optics that cause ray angles to be different. This section also shows that there are four kinds of lenses that are in common use: entocentric lenses (perspective in object and image space, with possibly differing ray angles), image-side telecentric lenses (perspective in object space, orthographic in image space), object-side telecentric (orthographic in object space, perspective in image space), and bilateral telecentric (orthographic in object and image space). Each of these kinds of lenses can be tilted, which obviously results in a different imaging geometry for each lens type.

A review of existing camera models for tilt lenses (cf. Sect. 5) has revealed several important shortcomings. We

---

Communicated by Stephen Lin.

---

✉ Carsten Steger  
steger@mvtec.com

<sup>1</sup> MVTec Software GmbH, Arnulfstraße 205, 80634 Munich, Germany

therefore propose a comprehensive camera model for tilt lenses, consisting of a camera model for each of the four kinds of possible tilt lenses, that removes these shortcomings (see Sect. 7 for the proposed models and Sect. 6 for the regular models on which the models we propose are based).

We will also examine under which conditions the parameters of the proposed models can be determined uniquely and under which conditions degeneracies occur (Sect. 7.2). Furthermore, we will show in Sect. 8 that all types of projective cameras can be regarded as tilt cameras in a natural manner.

The camera model we propose is versatile. We show in Sect. 12 that the model can also be used to describe the rectification geometry of a stereo image pair of a perspective and a telecentric camera.

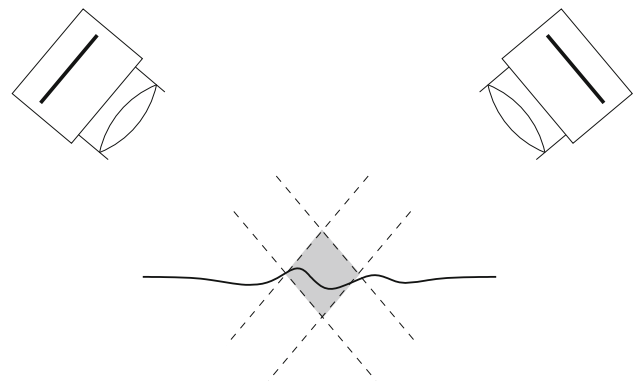
We also describe an algorithm that is capable of calibrating an arbitrary combination of perspective and telecentric cameras (Sect. 9). Furthermore, we describe two computationally efficient and accurate algorithms to remove the bias from circular control points (Sect. 10). Finally, we test the proposed camera models on many real lenses (Sect. 11).

## 2 Applications

In cameras that are equipped with regular lenses, the portion of the object space that is imaged in focus is parallel to the image plane. The depth of field, i.e., the range of object distances that appear focused in the image, is proportional to the inverse of the square of the magnification of the lens (Lenhardt 2006, Chapter 4.2.13.2). As usual, magnification is defined as the ratio of the size of the image of an object to the size of the object, where size denotes the distance to the optical axis (Lenhardt 2006, Chapter 4.2.3). Consequently, the larger the magnification of the camera, the smaller the depth of field is.

The small depth of field at high magnifications becomes problematic whenever it is necessary to image objects in focus that lie in or close to a plane that is not parallel to the image plane. With regular lenses, this is only possible by reducing the size of the aperture stop (typically, the diaphragm), i.e., by increasing the  $f$ -number of the lens. However, there is a limit to this approach for two reasons. First, if the aperture stop is made too small, the image will appear blurred because of diffraction (Lenhardt 2006, Chapters 4.3.4–4.3.7). Second, a small aperture stop causes less light to reach the sensor. Consequently, high-powered illumination is required to achieve reasonable exposure times, especially when images of moving objects must be acquired.

There are numerous practical applications in which a plane in object space that is not parallel to the image plane must be imaged in focus. One example is stereo reconstruction, where typically the cameras are used in a converging setup. As shown in Fig. 1, this setup causes the volume in object



**Fig. 1** A regular stereo setup with converging cameras. The image planes are visualized by *thick solid lines*. The depth of field of the two cameras is visualized by *dashed lines*. The common depth of field is visualized by the *gray rhombus*. The angles of the two cameras are exaggerated to display the common depth of field more clearly. The surface to be reconstructed is visualized by a *thick solid line*

space for which both cameras produce a sharp image to be a rhomboid-shaped infinite prism. This problem is typically ignored at small magnifications because the common depth of field is large enough. For large magnifications (e.g., larger than 0.1), however, the volume is small enough to cause significant defocus.

One method to cause the volumes in object space that are imaged in focus to become parallel is to use shift lenses, i.e., lenses that are shifted with respect to the sensor, while the optical axis remains perpendicular to the sensor (Willert 1997; Prasad 2000). One major problem is that this approach requires imaging obliquely through the lens, which increases vignetting and optical aberrations and thus decreases image quality (Willert 1997; Prasad 2000). In contrast, using the converging stereo configuration avoids these problems since the rays pass through the lens at angles close and symmetric to the optical axis, i.e., in the area for which lenses are designed to have little aberrations. Furthermore, shifting the lens is impossible if telecentric lenses are used in the stereo setup. Here, obviously, a converging geometry must be used to have any parallax in the images.

Another application where a tilted object plane must be imaged in focus is sheet-of-light 3D reconstruction. Here, a laser with a cylindrical lens projects a laser line onto objects in the world and a camera acquires images of the reflections of the laser line (Gross 2005, Chapter 16.6.1; Beyerer et al. 2016, Chapters 7.3.1 and 7.3.2; Li et al. 2007; Legarda et al. 2011). The projection of the laser line forms a plane in space that is not perpendicular to the optical axis of the camera. Different object distances cause different displacements of the laser line in the image, which allow a 3D reconstruction of the scene. To obtain maximum accuracy, it is necessary that the laser line is in focus for all 3D depths that must be reconstructed, i.e., the entire 3D laser plane emitted by

the projector should ideally be in focus. In particle image velocimetry applications, the laser sheet approach is combined with a stereo camera setup (Willert 1997; Prasad 2000). This allows the out-of-plane motion component of the velocity vector of the particles to be determined.

Another application where it is important to image a plane in focus that is tilted with respect to the image plane is fringe projection (Gross 2005, Chapter 16.6.2; Beyrer et al. 2016, Chapter 7.3.4; Albers et al. 2015; Peng et al. 2015). Here, a 2D projector replaces one of the cameras of a stereo camera setup. Consequently, this application is geometrically equivalent to the stereo camera setup described above.

### 3 Tilt Lenses and the Scheimpflug Principle

It is well known that an arbitrary plane in object space can be imaged in focus by tilting the lens with respect to the image plane (Gross 2005, Chapter 10.5.2). This principle is traditionally credited to Theodor Scheimpflug, who filed a series of Austrian, British, and US patents on the subject in 1902–1904. The patents were granted in 1904–1906 (Scheimpflug 1902a, b, c, 1903a, b, c, 1904). However, the principle was already known to Jules Carpentier, who filed for and received a British patent that describes the principle in 1901 (Carpentier 1901). To Scheimpflug’s credit, he worked out the optical and mathematical principles in detail in his patents, whereas Carpentier simply stated the Scheimpflug principle without any proof.

If a thin lens model is assumed, the Scheimpflug principle states the following:<sup>1</sup> The object plane (the plane that is in focus), the thin lens’s plane, and the image plane must all meet in a single line. This version of the Scheimpflug principle appears in (Carpentier 1901). The line of intersection is often called the Scheimpflug line (Evens 2008a, b; Merklinger 2010), while Scheimpflug calls it the axis of collineation. In his original Austrian patents, Scheimpflug uses a thick lens model,<sup>2</sup> for which the condition must be modified as follows: the Scheimpflug line is split into two lines, one in each principal plane of the lens, that are conjugate to each other, i.e., have the same distance and orientation with respect to the principal points of the lens (see Fig. 2). The angles of the object and image planes with respect to the principal planes can be derived from the lens equation and are given by (Scheimpflug 1904):

$$\tan \tau' = \frac{f}{a - f} \tan \tau, \tag{1}$$

<sup>1</sup> The optics terms in this section will be explained in Sect. 4.

<sup>2</sup> In his British and US patents, he also mentions the thin lens model.

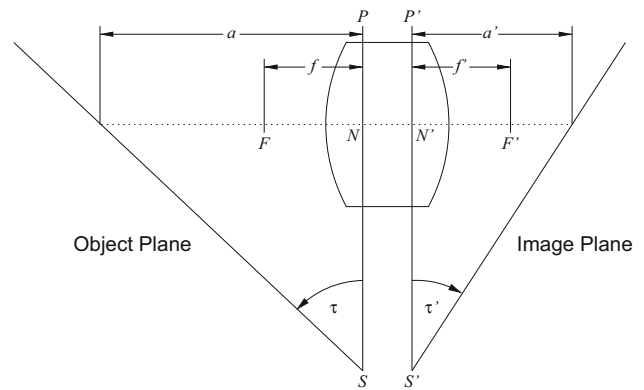


Fig. 2 The Scheimpflug principle.  $S$  and  $S'$  are the object-side and image-side Scheimpflug lines

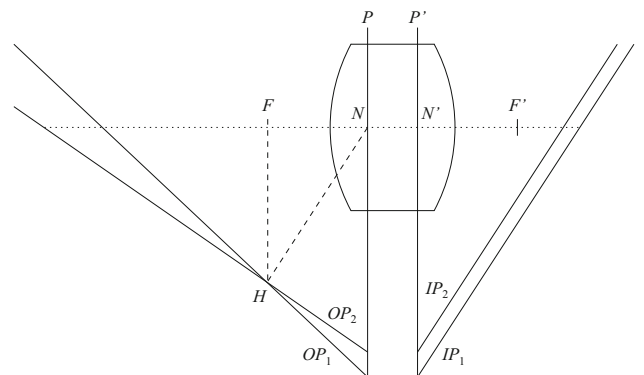
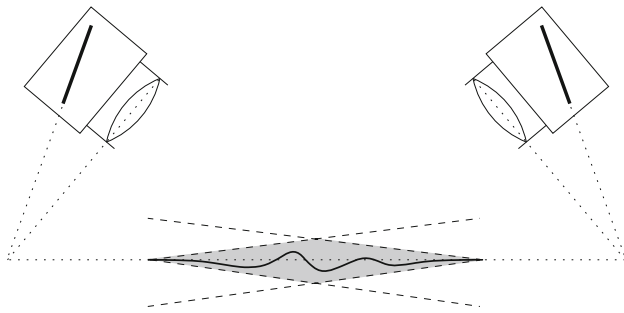


Fig. 3 Refocusing by moving the image plane with respect to the principal plane  $P'$  from  $IP_1$  to  $IP_2$  rotates the object plane that is in focus from  $OP_1$  to  $OP_2$  around the hinge line  $H$

where  $f$  is the focal length of the lens,  $\tau$  is the angle of the object plane with respect to the object-side principal plane,  $\tau'$  is the angle of the image plane with respect to the image-side principal plane, and  $a$  is the distance of the intersection point of the optical axis with the object plane from the object-side principal point.

If a plane that is parallel to the image plane is drawn through the object-side principal point, and this plane is intersected with the object-side focal plane, a straight line is obtained (see Fig. 3). This construction can also be performed analogously on the image side of the lens. Scheimpflug calls these lines the counter axes, while in current publications they are called hinge lines (Evens 2008a, b; Merklinger 2010) or pivot lines (Wheeler 2003). The object-side hinge line  $H$  has an important geometric significance: if the image is refocused (either by changing the distance of the image plane with respect to the image-side principal point or by changing the focus setting on the lens, which effectively moves the location of the principal planes), the object plane will rotate around the object-side hinge line if the tilt of the image plane remains fixed (Wheeler 2003; Evens 2008a, b; Merklinger 2010). Furthermore, the depth of field is an infinite wedge-

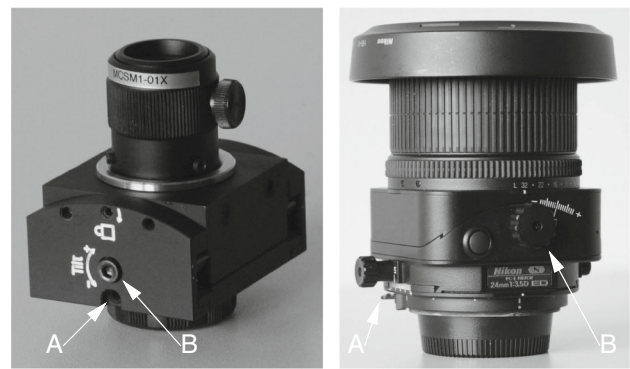


**Fig. 4** A stereo setup with converging cameras and image planes tilted according to the Scheimpflug principle. The image planes are visualized by *thick solid lines*. The depth of field of the two cameras is visualized by *dashed lines*, which emanate from the hinge lines. The common depth of field is visualized by the *gray rhombus*. The angles of the two cameras are exaggerated to display the common depth of field more clearly. The surface to be reconstructed is visualized by a *thick solid line*

shaped region that has the hinge line as its edge (Wheeler 2003; Evens 2008a, b; Merklinger 2010). This can also be seen from Fig. 3. If we interpret the image planes  $IP_1$  and  $IP_2$  as the limits of the depth of focus, the limits of the depth of field are given by  $OP_1$  and  $OP_2$ .<sup>3</sup> Note that positioning the image plane parallel to the principal planes moves the hinge and Scheimpflug lines to infinity, and produces the regular depth-of-field geometry.

With the Scheimpflug principle, it is obvious how to solve the focusing problems in the above applications. For example, for stereo cameras, the image planes must be tilted as shown in Fig. 4. A similar principle holds for fringe projection systems. For sheet-of-light systems, the object plane in Fig. 4 corresponds to the laser plane and there is only one camera.

To construct a camera with a tilted lens, there are several possibilities. One popular option is to construct a special camera housing with the desired tilt angle, to which the lens can be attached. This is typically done for specialized applications for which the effort of constructing the housing is justified, e.g., specialized sheet-of-light or fringe-projection sensors. In these applications, the lens is typically tilted around the vertical or horizontal axis of the image, as required by the application. Another possibility is to use a Scheimpflug adapter that allows an arbitrary lens to be attached to a regular camera. Again, often only horizontal or vertical tilt is supported. The most versatile option is to use lenses that have been designed specifically to be tilted in an arbitrary direction. In machine vision parlance, these lenses are typically called Scheimpflug lenses or Scheimpflug optics. They are available as perspective or telecentric lenses. In the con-



**Fig. 5** Examples for a machine vision tilt lens (*left*) and an SLR tilt lens (*right*). In both cases, the *letter A* denotes a lock mechanism that can be unlocked (a screw for the machine vision lens and a lever for the SLR lens). After unlocking, the upper part of the lens can be rotated with respect to the part of the lens that is attached to the camera housing to determine the direction into which the lens is tilted. The *letter B* denotes the mechanism by which the lens can be tilted (a screw for the machine vision lens and a knob for the SLR lens)

sumer SLR camera market, these lenses are typically called tilt/shift lenses or perspective correction lenses (although, technically, perspective correction only requires that the lens can be shifted). Since the ability to tilt the lens is the important feature for the purposes of this paper, we will call these lenses tilt lenses from now on.

Figure 5 shows examples for a machine vision and an SLR tilt lens. In both cases, one first selects the direction into which the lens should be tilted by rotating the entire lens assembly, except the part that is attached to the camera housing, and then tilts the lens into that direction. Ideally, a camera model would represent these two tilt parameters in a natural manner that is easy to understand for the user.

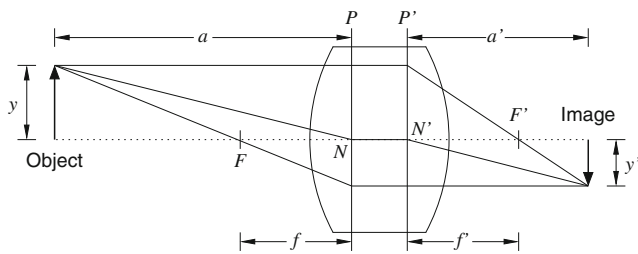
## 4 Required Principles of Geometric Optics

The pinhole model is in widespread use in machine and computer vision to model the projection that standard perspective lenses perform. It is a convenient model that is an abstraction of the projection that happens in a real lens. To develop a projection model for tilt lenses, the pinhole model is inadequate, as we will discuss in this section.

It is well known from Gaussian optics (Lenhardt 2006, Chapter 4.2.11) that any assembly of lenses can be regarded as a single thick lens. The thick lens model is characterized by certain cardinal elements, whose main purpose is to simplify ray tracing through the lens (Lenhardt 2006, Chapter 4.2.4). The principal planes  $P$  and  $P'$  are perpendicular to the optical axis. A ray that enters the object-side principal plane  $P$  exits the image-side principal plane  $P'$  at the same position with respect to the optical axis (however, in general not at the

<sup>3</sup> This is an approximation that is accurate enough for most applications. The actual shape of the depth of field is slightly more complicated, as discussed by Evens (2008b).





**Fig. 6** The geometry of a thick lens

same angle).<sup>4</sup> The principal points are the intersections of the principal planes with the optical axis. The focal points  $F$  and  $F'$  lie on the optical axis at distances  $f$  and  $f'$  from  $P$  and  $P'$ . For lenses that have the same medium, e.g., air, on both sides of the lens,  $f$  and  $f'$  have the same absolute value.<sup>5</sup> A ray that is parallel to the optical axis in image space or object space passes through  $F$  or  $F'$ , respectively. Finally, the nodal points  $N$  and  $N'$  lie on the optical axis and have the property that a ray that passes through one nodal point emerges at the other nodal point with the same direction. For lenses that have the same medium on both sides of the lens, the nodal points coincide with the principal points.

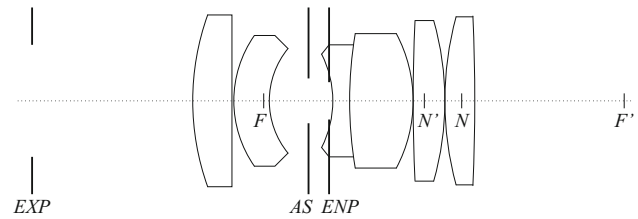
Figure 6 displays a thick lens with its cardinal elements and three rays that show the three most important principles derived from the above properties (Lenhardt 2006, Chapter 4.2.4):

- A ray that is parallel to the optical axis in object space passes through the image-side focal point.
- A ray that is parallel to the optical axis in image space passes through the object-side focal point.
- A ray that passes through a nodal point exits at the other nodal point with unchanged direction.

If the thick lens model is simplified to the thin lens model, the two principal planes coincide and the two nodal points collapse to the center of the thin lens (Lenhardt 2006, Chapter 4.2.5). While the above three rules remain valid, they simplify. In particular, the third rule simplifies to: a ray that passes through the center of a thin lens does not change its direction. At first glance, this appears to be very similar to the ray geometry of a pinhole camera. It might, therefore, seem obvious to define the projection center of a thin lens as the center of the lens (see, e.g., Hanning 2011, Chapter 2.2). Similarly, for a thick lens, it might seem obvious to define that there are two projection centers: the two nodal points (see, e.g., Mikhail et al. 2001, Chapter 4.3). These defini-

<sup>4</sup> Note that  $P$  may be closer to the image plane than  $P'$ .

<sup>5</sup> Various sign conventions are in common use. For example, according to DIN 1335:2003-12 (2003),  $f$  is negative and  $f'$  positive. In this paper, we use the convention that both  $f$  and  $f'$  are positive.



**Fig. 7** The pupils and cardinal elements of a real system of lenses. AS is the aperture stop (a diaphragm); ENP is the entrance pupil; EXP is the exit pupil

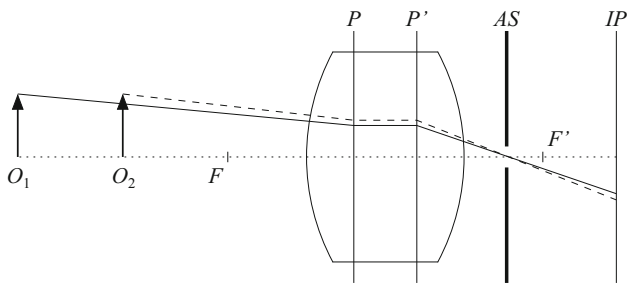
tions mainly would be motivated by the desire to preserve the property of pinhole cameras that the ray directions are identical on both sides of the pinhole.

It is, however, incorrect to regard a thin or thick lens as a pinhole camera. The projection properties of a lens are determined by one essential element of the lens that is neglected by both the thin and thick lens models: the aperture stop. If an object is imaged through a lens, a cone of rays is emitted from the object, passes through the lens, and is imaged sharply in the image plane. The aperture stop is the opening in the lens that restricts the diameter of the cone of rays (Lenhardt 2006, Chapter 4.2.12.2). Thus, it is the analog of the pinhole in a pinhole camera. The virtual image of the aperture stop by all optical elements that come before it (i.e., lie on the object side of the aperture stop) is called the entrance pupil, while the virtual image of the aperture stop by all optical elements that come behind it (i.e., lie on the image side of the aperture stop) is called the exit pupil. Figure 7 shows the cardinal elements of a real system of lenses along with the aperture stop (diaphragm) and the two pupils.

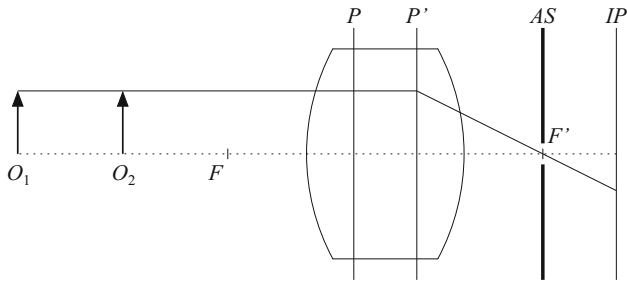
The cone of rays emitted by the object seems to pass through the entrance pupil and seems to emerge from the exit pupil. A ray that passes through the center of the aperture stop is called a chief ray. Since it lies in the center of the cone of rays (with respect to the plane of the aperture stop), it is the analog of the ray through the pinhole of a pinhole camera (Lenhardt 2006, Chapter 4.2.12.4). Note that the chief ray also passes through the center of the entrance and exit pupils (because they are the virtual images of the aperture stop).

The placement of the aperture stop is crucial for the projection properties of a lens. The aperture stop essentially acts as a filter that filters out certain rays in object space.

Often, the aperture stop is placed between the two focal points of the lens, resulting in an entocentric lens (Lenhardt 2006, Chapter 4.2.14.1), as shown in Fig. 8. In entocentric lenses, the entrance and exit pupils lie at finite positions and between the object and the image plane. One important point to note here is that the angle of the chief ray in object space in general is different from the angle of the chief ray in image space (Lenhardt 2006, Chapter 4.2.12.3). Because of the third rule above, the angles would only be identical if the aperture stop were located in one of the principal planes.



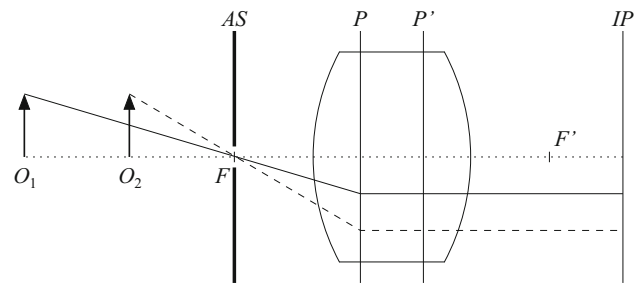
**Fig. 8** An entocentric lens.  $O_1$  and  $O_2$  are two different object positions;  $AS$  is the aperture stop;  $IP$  is the image plane. For simplicity, only the chief rays are displayed. The image of  $O_1$  is in focus, while the image of  $O_2$  is blurred. Note that the image of  $O_2$  is larger than the image of  $O_1$



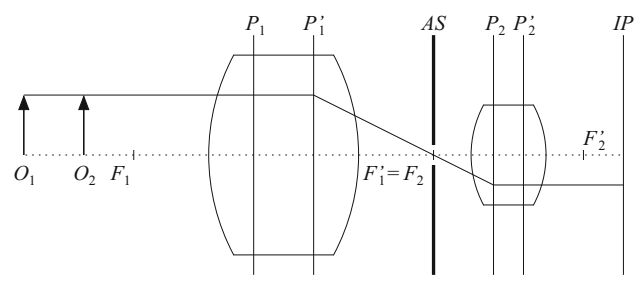
**Fig. 9** An object-side telecentric lens.  $O_1$  and  $O_2$  are two different object positions;  $AS$  is the aperture stop;  $IP$  is the image plane. For simplicity, only the chief rays are displayed. The image of  $O_1$  is in focus, while the image of  $O_2$  is blurred. Note that the images of  $O_1$  and  $O_2$  have the same size

The aperture stop can also be placed at one of the focal points. If an infinitely small aperture stop is positioned at the image-side focal point, according to the first rule above, only rays parallel to the optical axis in object space can pass the aperture stop, i.e., the lens performs a parallel projection in object space (Lenhardt 2006, Chapter 4.2.14.1). The entrance pupil is the virtual image of the aperture stop and is therefore located at infinity (the exit pupil is at a finite location; in this simplified model, it lies at the aperture stop; in a real lens, there could be some optical elements behind the aperture stop). Because the entrance pupil lies at infinity, these lenses are called object-side telecentric. In real lenses, the aperture stop has finite size to allow enough light to reach the sensor. This does not change the telecentric properties (Lenhardt 2006, Chapter 4.2.14.2). However, the chief ray is replaced by a cone of rays that has the chief ray as its center. This causes a finite depth-of-focus and depth-of-field. An object-side telecentric lens is shown in Fig. 9.

In an analogous manner, if the aperture stop is positioned at the object-side focal point, by the second rule above, only rays that are parallel to the optical axis in image space can pass the aperture stop. The entrance pupil remains finite, while the exit pupil moves to infinity. These kinds of lenses are called image-side telecentric. Lenses that are designed for



**Fig. 10** An image-side telecentric lens.  $O_1$  and  $O_2$  are two different object positions;  $AS$  is the aperture stop;  $IP$  is the image plane. For simplicity, only the chief rays are displayed. The image of  $O_1$  is in focus, while the image of  $O_2$  is blurred. Note that the image of  $O_2$  is larger than the image of  $O_1$

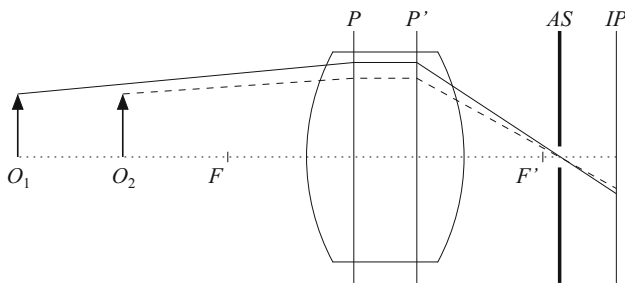


**Fig. 11** A bilateral telecentric lens.  $O_1$  and  $O_2$  are two different object positions;  $AS$  is the aperture stop;  $IP$  is the image plane. For simplicity, only the chief rays are displayed. The image of  $O_1$  is in focus, while the image of  $O_2$  is blurred. Note that the images of  $O_1$  and  $O_2$  have the same size

digital sensors are often image-side telecentric or nearly so to avoid pixel vignetting (the effect that solid-state sensors are less sensitive to rays that impinge on the sensor non-perpendicularly). An image-side telecentric lens is shown in Fig. 10.

If an object-side telecentric lens is modified by placing a second lens assembly behind the image-side focal point in such a way that the object-side focal point of the second lens coincides with the image-side focal point of the first lens, i.e., if the object-side focal point of the second lens lies in the center of the aperture stop of the first lens, we obtain a bilateral telecentric lens (Lenhardt 2006, Chapter 4.2.14.2). This construction essentially is identical to attaching an image-side telecentric lens to an object-side telecentric lens (with a single aperture stop, of course). In bilateral telecentric lenses, both the entrance and exit pupils are located at infinity. The advantage of bilateral telecentric lenses over object-side telecentric lenses is that the telecentricity is achieved more accurately since the circle of confusion for bilateral telecentric lenses is more symmetric and suffers less from lens aberrations than for object-side telecentric lenses. Figure 11 shows the ray geometry of a bilateral telecentric lens.

As a final example, the aperture stop can be placed behind the image-side focal point, as shown in Fig. 12. This causes

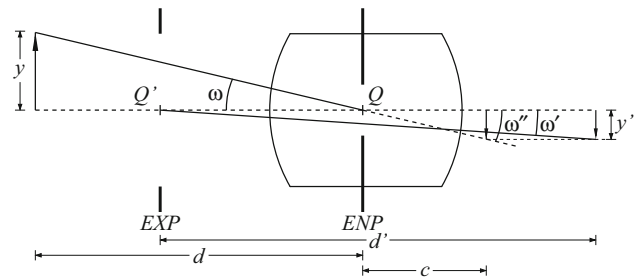


**Fig. 12** A hypercentric lens.  $O_1$  and  $O_2$  are two different object positions;  $AS$  is the aperture stop;  $IP$  is the image plane. For simplicity, only the chief rays are displayed. The image of  $O_1$  is in focus, while the image of  $O_2$  is blurred. Note that the image of  $O_2$  is smaller than the image of  $O_1$

the entrance pupil to lie in object space. If an object is placed between the entrance pupil and the lens, this configuration has the effect that objects that are closer to the camera appear smaller in the image. The canonical example is an image of a die that is oriented parallel to the image plane. This lens geometry enables five of the six sides of the die to be seen in the image. These types of lenses are typically called hypercentric lenses (Lenhardt 2006, Chapter 4.2.14.1; Gross 2005, Chapter 10.3.9; Beyerer et al. 2016, Chapter 3.4.6).

As these examples have shown, simply moving the aperture stop completely changes the projection properties of the lens, even if all the cardinal elements remain unchanged. This shows that neither the nodal points (in the thick lens model) nor the center of the lens (in the thin lens model) can be the projection centers of the lens. The actual projection centers of the lens that explain the projection behaviors that were described above are the centers of the entrance and exit pupils (Lenhardt 2006, Chapter 4.2.12.4; Luhmann et al. 2014, Chapter 3.3.2.2; Evens 2008b, Section 10). In particular, the center of the entrance pupil is the relevant projection center for many computer and machine vision applications since it determines the ray geometry in object space. For example, the center of the entrance pupil is the point around which a camera must be rotated to create a panorama image (Evens 2008b, Section 10).

Note that all the lens models described above are very different from the pinhole model. First, there are two projection centers. Second, in general the ray angles in image space and object space are different. This is the case even for entocentric lenses. An explanation of how the pinhole model is connected to entocentric lenses is given by Lenhardt (2006, Section 4.2.12.4) and Steger et al. (2008, Chapter 2.2.2). To map the entocentric lens model to the pinhole model, we must ensure that the object and image side ray angles are identical. All the ray angles  $\omega$  in object space must remain fixed since they are determined by the geometry of the objects in the scene. Therefore, the center of the entrance pupil must remain fixed. We can, however, move the exit pupil to the



**Fig. 13** Modeling an entocentric lens as a pinhole lens.  $ENP$  and  $EXP$  are the entrance and exit pupils with centers  $Q$  and  $Q'$ . The chief ray enters the entrance pupil at an angle  $\omega$  and exits from the exit pupil at a different angle  $\omega'$ . An object of size  $y$  is projected to an image of size  $y'$ . The distance from the object to the entrance pupil is  $d$ , the distance from the exit pupil to the image is  $d'$ . To simulate a pinhole lens,  $Q'$  virtually must be moved to  $Q$ . To make the image-side ray angle  $\omega'$  identical to  $\omega$ , the image plane must be moved to a distance  $c$  (the principal distance) from  $Q$ , while keeping  $y'$  fixed. Note that  $c$  and  $d'$  differ substantially

same position as the entrance pupil. Finally, if the image plane is perpendicular to the optical axis, we can move the image plane to a position where the image-side ray angles  $\omega'$  become identical to the object-side ray angles  $\omega$ , while keeping the image size  $y'$  fixed. This is shown in Fig. 13. Note that this construction even works for image-side telecentric lenses.

For lenses that are telecentric in object space (object-side and bilateral telecentric lenses), we can also pretend that the image space rays have the same angle as the object space rays if the image plane is perpendicular to the optical axis. For bilateral telecentric lenses, this is tautologically true. For object-side telecentric lenses, we can move the image plane to infinity to achieve this. Hence, for both lens types, we can assume that the camera performs an orthographic projection.

If the image plane is tilted with respect to the optical axis, these convenient abstractions unfortunately no longer work correctly. From the examples above, it is obvious that tilting the image plane causes perspective or affine distortions of the image. In this case, it is essential to be able to model the ray geometry in image space correctly to obtain the correct perspective. Let us imagine that the camera is acquiring an image of a square that is perpendicular to the optical axis with a tilted image plane. For an entocentric lens, the image of the square will be a trapezoid.<sup>6</sup> However, if we do not model the ray geometry in image space correctly, the trapezoid will have the wrong shape.<sup>7</sup> If we use an image-side telecentric lens, the image of the square will be a rectangle.

<sup>6</sup> This assumes that the tilt is in the direction of one of the axes of the square. If not, a general quadrilateral will result.

<sup>7</sup> This is clear from Fig. 13. If we imagine the image plane tilted by the same angle, the fact that  $d'$  and  $c$  differ will cause different perspective distortions that are similar to acquiring images of a tilted plane with different focal lengths with a regular lens, while keeping the size of the

Therefore, we can see that while the pinhole model is able to model entocentric and image-side telecentric lenses correctly if the image plane is perpendicular to the optical axis, this clearly is no longer the case for tilt lenses (tilted image planes). Similarly, for bilateral telecentric lenses, the image of the square is a rectangle, while for object-side telecentric lenses, it is a trapezoid. Again, it is evident that the orthographic projection model is inadequate for tilt lenses. From this discussion, it is clear that the projection center that is relevant for modeling the tilt correctly is the center of the exit pupil and its geometric relation to the image plane.

We close this section by mentioning that the ray angles  $\omega$  and  $\omega'$  (cf. Fig. 13) are related to the pupil magnification factor  $m_p$  by (see [Lenhardt 2006](#), Chapter 4.2.12.3)

$$\frac{\tan \omega}{\tan \omega'} = m_p, \quad (2)$$

where  $m_p$  is the diameter of the exit pupil divided by the diameter of the entrance pupil. Consequently, whenever the entrance and exit pupils have different sizes, the ray angles in object and image space will be different. Therefore, by simply looking at the lens from the front and back, we can easily determine whether the ray angles are different.

## 5 Related Work

### 5.1 Related Work on the Calibration of Cameras with Tilt Lenses

In this section, we will discuss previous approaches to the calibration of cameras with tilt lenses. We will only discuss approaches that provide an explicit geometric camera model and that calibrate the model parameters. For example, we will not discuss approaches that only compute a rectification from the image plane to the laser plane ([Willert 1997](#); [Walker 2002](#); [Konrath and Schröder 2000, 2002](#)), or that do not describe how the camera parameters are calibrated ([Fournel et al. 2003](#)).

A review of existing approaches has shown that there are problems that occur frequently. To avoid tedious repetitions, rather than discussing the existing approaches individually, we discuss the problems and list the approaches together with the problems. Every existing approach discussed below exhibits at least two of the problems.

It is interesting to note that there is no single approach that can handle all the lens types that are relevant for machine vision applications (perspective, image-side telecentric, object-side telecentric, and bilateral telecentric, as

discussed in Sect. 4). Existing approaches always deal with a single lens type. Perspective lenses are treated in ([Gerdes et al. 1993a,b](#); [Gennery 2001, 2006](#); [Fournel et al. 2004, 2006](#); [Louhichi et al. 2006, 2007](#); [Haig et al. 2006](#); [Li et al. 2007](#); [Wang et al. 2008](#); [Legarda et al. 2011, 2013](#); [Hamrouni et al. 2012](#); [Astarita 2012](#); [Cornic et al. 2015](#); [Fasogbon et al. 2015](#); [Albers et al. 2015](#); [Kumar and Ahuja 2014a,b, 2015](#)) and object-side telecentric lenses in ([Peng et al. 2015](#)). There are no approaches for image-side and bilateral telecentric lenses.

As discussed in Sect. 3, a lens can typically be tilted in an arbitrary direction (see also [Walker 2002](#)). Yet, there are approaches that only model a lens tilt in one direction. For some approaches ([Fournel et al. 2006](#); [Louhichi et al. 2006, 2007](#); [Astarita 2012](#)), it is unclear in which direction the tilt is modeled, but there is only a single tilt parameter, which precludes tilting in an arbitrary direction. Furthermore, [Li et al. \(2007\)](#) only model a tilt around the x-axis, while [Haig et al. \(2006\)](#) only model a tilt in the direction of gravity.

More importantly, almost all of the approaches that handle perspective lenses do not model different ray angles in object and image space ([Gerdes et al. 1993a,b](#); [Gennery 2001, 2006](#); [Fournel et al. 2004, 2006](#); [Louhichi et al. 2006, 2007](#); [Haig et al. 2006](#); [Li et al. 2007](#); [Wang et al. 2008](#); [Legarda et al. 2011, 2013](#); [Hamrouni et al. 2012](#); [Astarita 2012](#); [Kumar and Ahuja 2014b](#); [Cornic et al. 2015](#); [Fasogbon et al. 2015](#); [Albers et al. 2015](#)). The only exception is the approach by [Kumar and Ahuja \(2014a, 2015\)](#).<sup>8</sup> Their camera model, however, is unnecessarily complex and contains redundant parameters. In their approach, the tilt and differing ray angles are modeled by at least four parameters, whereas three are sufficient, as will be described in Sect. 7. Furthermore, the model requires data to be specified that few lens manufacturers provide, such as the distances from the two principal planes to the respective pupils. This makes it impractical to use for typical machine vision users.

As will be shown in this paper, the model of the mapping of tilted lenses includes important degeneracies that imply that certain parameters of the model cannot be determined uniquely. These degeneracies are not analyzed by any existing approach ([Gerdes et al. 1993a,b](#); [Gennery 2001, 2006](#); [Fournel et al. 2004, 2006](#); [Louhichi et al. 2006, 2007](#); [Haig et al. 2006](#); [Li et al. 2007](#); [Wang et al. 2008](#); [Legarda et al. 2011, 2013](#); [Hamrouni et al. 2012](#); [Astarita 2012](#); [Cornic](#)

<sup>8</sup> The work by [Kumar and Ahuja \(2014a, 2015\)](#) is based on the work by [Aggarwal and Ahuja \(2002\)](#), which points out some anomalies that can happen if different ray angles in object and image space are not modeled correctly. They only propose a calibration for the subset of parameters that model the different ray angles in object and image space, i.e., they do not describe a complete calibration of tilt cameras. The calibration requires a complex and elaborate setup. Since this makes it impractical to use for typical machine vision users, we do not include it in the above discussion.

Footnote7 continued

plane identical in the image (see, e.g., [Faugeras and Luong 2001](#), Figure 4.21 or [Hartley and Zisserman 2003](#), Figure 6.7).



et al. 2015; Fasogbon et al. 2015; Albers et al. 2015; Peng et al. 2015; Kumar and Ahuja 2014a, b, 2015).

The failure to model the different ray directions in object and image space together with the failure to exclude certain interior orientation parameters in degenerate cases, which would be necessary to handle the degenerate cases correctly, as shown in Sect. 7, means that most existing camera models can only handle the case of horizontal and vertical tilts without excessive residual errors (Gerdes et al. 1993a, b; Gennery 2001, 2006; Fournel et al. 2004, 2006; Louhichi et al. 2006, 2007; Haig et al. 2006; Li et al. 2007; Wang et al. 2008; Legarda et al. 2011, 2013; Hamrouni et al. 2012; Astarita 2012; Kumar and Ahuja 2014b; Cornic et al. 2015; Fasogbon et al. 2015; Albers et al. 2015). However, in this case the camera parameters will not correspond to the correct values, even if the model returns small residual errors.

Another problem is that many approaches parameterize the tilt mapping by Euler angles (Gerdes et al. 1993a, b; Wang et al. 2008; Legarda et al. 2011, 2013; Hamrouni et al. 2012; Cornic et al. 2015; Fasogbon et al. 2015; Albers et al. 2015; Peng et al. 2015; Kumar and Ahuja 2014a, b, 2015). As can be seen from the example lenses in Sect. 3, this kind of parameterization is unintuitive for the user if the tilt is not horizontal or vertical.

A few of the proposed approaches use a self-calibration method in which the coordinates of the control points of the calibration object are determined along with the camera parameters. It is well known that self-calibration cannot recover the scale of the scene, i.e., of the calibration object (Hartley and Zisserman 2003, Chapters 10.2 and 19.1; Luhmann et al. 2014, Chapter 4.4). This fact is ignored in (Fournel et al. 2006; Louhichi et al. 2006; Hamrouni et al. 2012). It is acknowledged implicitly in (Louhichi et al. 2007) or explicitly in (Fournel et al. 2004; Cornic et al. 2015), but no calibration for the undetermined scale is provided. However, it is known from photogrammetry that some form of measured distances or control points, e.g., a scale bar, are necessary to recover the scale (Luhmann et al. 2014, Chapters 4.4.3, 7.3.1.5, and 7.3.2.1; Haig et al. 2006). Known control points or distances effectively constitute a calibration object. Therefore, we might as well use a calibration object with known control points.

Some of the calibration approaches use circular control points. It is well known that this may lead to biased calibration results if the bias is not taken into account and removed (Heikkilä 2000). Many of the existing approaches exhibit this problem (Gennery et al. 1987; Gennery 2001, 2006; Fournel et al. 2004; Louhichi et al. 2006, 2007; Li et al. 2007; Haig et al. 2006; Legarda et al. 2011; Astarita 2012; Cornic et al. 2015; Albers et al. 2015; Peng et al. 2015). In some approaches, the bias is taken into account (Legarda et al. 2013; Fasogbon et al. 2015), albeit in an inefficient manner

by the methods of Datta et al. (2009) or Vo et al. (2011), to be discussed in Sect. 5.2.

Finally, some approaches inherently require a 3D calibration object (Gerdes et al. 1993a, b; Gennery 2001, 2006). This is a disadvantage in many machine vision applications because 3D calibration objects are more expensive to manufacture accurately and they are sometimes more cumbersome to handle for the users.

## 5.2 Related Work on the Bias Removal for Circular Control Points

It is well known that there are two kinds of bias that may affect the location of the projection of circular control points in the image. First, in an ideal pinhole camera that does not have lens distortions, the projection of the circular control point into the image is an ellipse. The center of the ellipse is typically used as the projected center of the corresponding control point. However, the projection of the center of a circle in 3D in general is not the center of the ellipse in the image (Lenz 1988, Chapter 10.5; Lenz and Fritsch 1990; Ahn et al. 1999; Heikkilä 2000; Mallon and Whelan 2007). Therefore, using the centers of the ellipses introduces a bias in the control point locations in the image that affects the accuracy of the camera calibration. Like Mallon and Whelan (2007), we will call this bias the perspective bias.

Second, if the lens exhibits lens distortions, the projection of a circle in 3D into the image typically is no longer an ellipse. Thus, if the camera calibration assumes an ellipse in the image to perform the extraction of control points, e.g., by fitting an ellipse to the control point's edges in the image (Heikkilä 1997, Chapter 2; Lanser 1997, Chapter 3.2.1.1) or by using gray value moments (Mallon and Whelan 2007), there will be a further bias that can affect the accuracy of the calibration (Lenz 1988, Chapter 10.5; Lenz and Fritsch 1990; Mallon and Whelan 2007). Like Mallon and Whelan (2007), we will call this bias the distortion bias.

The simplest approach to deal with the bias is to ignore it. As discussed in Sect. 5.1, this is a popular approach.

A slightly more elaborate approach is to require that the circles of the control points must be small enough that the perspective bias (and, consequently, the distortion bias) is sufficiently small (Dold 1996, 1997; Ahn et al. 1999; Mallon and Whelan 2007). Furthermore, Dold (1997) and Ahn et al. (1999) claim that, while the perspective bias negatively affects the interior and exterior orientation of the cameras in the calibrated setup, a 3D reconstruction obtained from the biased orientation parameters is largely unaffected by the bias. It was later shown by Otepka (2004) that this assertion is not true in general.

The approaches by Fournel et al. (2006) and Hamrouni et al. (2012), mentioned in Sect. 5.1, use the intersection points of bitangents to the circular marks as control points.

While these points are invariant to projection, they are not invariant to lens distortion, i.e., some bias remains.

An elegant solution to remove the perspective bias was proposed in (Heikkilä 1997, Chapter 3.5; Heikkilä 2000). The camera is first calibrated using the biased ellipse centers. Then, the calibrated camera parameters are used to determine the ellipse centers by projecting the control point borders into the image and determining their centers analytically. Furthermore, the projection of the center of the control point is determined analytically. These two points define an offset that is used to correct the extracted ellipse center and then to perform the calibration again using the thus corrected control points. For self-calibration, a similar approach is proposed by (Otepka 2004; Otepka and Fraser 2004). Since the distortion bias is not handled by these approaches, it can be expected that some bias remains in the calibrated camera parameters.

We now turn to approaches that address the perspective as well as the distortion bias.

The effect of perspective and distortion bias is analyzed in (Lenz 1988, Chapter 10.5; Lenz and Fritsch 1990). Approximate formulas (based on several simplifications of the problem) are given that allow to predict the two kinds of bias if the interior and exterior orientation of the camera are known. Lenz (1988, Chapter 10.5) describes an approach to reduce the bias. The camera is first calibrated with the biased control point locations. Then, the approximations to the two kinds of bias are predicted and are applied to the control point locations. No analysis of how the accuracy of the camera parameters is improved by this approach is performed. Since the two kinds of bias are treated only in an approximate manner, it can be assumed that some residual bias is still present in the camera parameters.

The approach by Douxchamps and Chihara (2009) follows the same general steps that the above approaches use (iteration of the camera calibration and the improvement of the control point positions). The improvement of the control point locations is performed by computing a synthetic template of the control point based on the current orientation parameters and then maximizing the correlation of the template and the image around the current estimate of the control point location. The computation of the synthetic template is based on a kind of ray tracing algorithm and therefore is computationally very expensive. This makes the approach unattractive for time-critical applications.

Finally, the approaches by Datta et al. (2009) and Vo et al. (2011) also follow the general scheme of iterating the camera calibration and the improving the control point positions. The difference is that the control point correction is performed by removing the lens distortions from each calibration image and projecting it into a plane that is parallel to the image plane using the current interior and exterior orientation parameters. The positions of the control points are then extracted in the rectified images and are projected into the image. The control

points are extracted either through ellipse fitting (Datta et al. 2009) or through template matching (Datta et al. 2009; Vo et al. 2011). Since these approaches rely on rectifying all calibration images once per outer iteration and then performing the control point extraction anew for each rectified image, they are very time-consuming, and therefore unattractive for time-critical applications.

## 6 Models for Cameras with Regular Lenses

### 6.1 Camera Models

The models for cameras with tilt lenses that will be proposed in Sect. 7 are all extensions of models for cameras with regular lenses. Therefore, we will discuss these models first. Our presentation is based on the description in (Steger et al. 2008, Chapter 3.9), but extends that description to the multi-view stereo case.

Suppose the multi-view stereo setup consists of  $n_c$  cameras. In the calibration, we use  $n_o$  images of a calibration object in different poses and acquire images of the calibration object with  $n_c$  cameras. The calibration object has its own coordinate system. Each pose  $l$  ( $l = 1, \dots, n_o$ ) of the calibration object defines a transformation from the calibration object coordinate system to a camera coordinate system of camera  $k$  ( $k = 1, \dots, n_c$ ). In applications, often a certain pose of the calibration object is used to define a world coordinate system (Steger 1998, Chapter 3.9.4)

First, a point  $\mathbf{p}_o = (x_o, y_o, z_o)^\top$  given in the calibration object coordinate system at pose  $l$  is transformed into a point  $\mathbf{p}_l$  in some reference camera coordinate system using a rigid 3D transformation:

$$\mathbf{p}_l = \mathbf{R}_l \mathbf{p}_o + \mathbf{t}_l, \quad (3)$$

where  $\mathbf{t}_l = (t_{l,x}, t_{l,y}, t_{l,z})^\top$  is a translation vector and  $\mathbf{R}_l$  is a rotation matrix that is parameterized by Euler angles:  $\mathbf{R}_l = \mathbf{R}_x(\alpha_l) \mathbf{R}_y(\beta_l) \mathbf{R}_z(\gamma_l)$ . Without loss of generality, we can assume that the reference camera coordinate system is that of camera 1. As usual, if we represent the points by homogeneous coordinates, we can write the transformation in (3) as a  $4 \times 4$  homogeneous matrix:

$$\mathbf{p}_l = \mathbf{H}_l \mathbf{p}_o = \begin{pmatrix} \mathbf{R}_l & \mathbf{t}_l \\ \mathbf{0}^\top & 1 \end{pmatrix} \mathbf{p}_o. \quad (4)$$

Next, the point  $\mathbf{p}_l$  is transformed into the camera coordinate system of camera  $k$  using

$$\mathbf{p}_k = \mathbf{R}_k \mathbf{p}_l + \mathbf{t}_k, \quad (5)$$

where  $\mathbf{R}_k = \mathbf{R}_x(\alpha_k)\mathbf{R}_y(\beta_k)\mathbf{R}_z(\gamma_k)$  and  $\mathbf{t}_k = (t_{k,x}, t_{k,y}, t_{k,z})^\top$  describe the relative pose of camera  $k$  with respect to camera 1. Again, we can represent this transformation by a homogeneous matrix  $\mathbf{H}_k$ . Note that, with the above convention,  $\mathbf{R}_k = \mathbf{I}$  and  $\mathbf{t}_k = \mathbf{0}$  for  $k = 1$ . Thus, if we only calibrate a single camera, the transformation in (5) is redundant and can be omitted. We will make use of this fact below to simplify the discussion where appropriate.

As discussed in Sect. 4, the relevant projection center for modeling the pose of the camera is the center of the entrance pupil. Therefore, we assume that the origin of the camera coordinate system is the center of the entrance pupil of the respective camera and that the  $z$  axis of the camera coordinate system points along the optical axis towards the scene.

Next, the point  $\mathbf{p}_k = (x_k, y_k, z_k)^\top$  is projected into the image plane. For lenses that are perspective in object space (i.e., perspective and image-side telecentric lenses), the projection is given by:

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \frac{c}{z_k} \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \tag{6}$$

where  $c$  is the principal distance. Note that we use the assumption that the ray angles in image space and object space are identical, which is a valid assumption since in this section we model the image plane as perpendicular to the optical axis (see the discussion in Sect. 4). If we use homogeneous coordinates, we can represent the projection in (6) by the  $3 \times 4$  matrix

$$\mathbf{P}_p = \begin{pmatrix} c & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{7}$$

For lenses that are telecentric in object space (i.e., object-side and bilateral telecentric lenses), the projection is given by:

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = m \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \tag{8}$$

where  $m$  is the magnification of the lens. Again, we use the assumption that the ray angles in image space and object space are identical. If we use homogeneous coordinates, we can represent the projection in (8) by the  $3 \times 4$  matrix

$$\mathbf{P}_o = \begin{pmatrix} m & 0 & 0 & 0 \\ 0 & m & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{9}$$

Conceptually, the undistorted point  $(x_u, y_u)^\top$  is then distorted to a point  $(x_d, y_d)^\top$ . In our approach, two distortion models are supported. The first model has been called the division model by Fitzgibbon (2001). It was in common use,

however, more than a decade before the term division model was coined (Lenz 1987, 1988; Lenz and Fritsch 1990; Lanser et al. 1995; Lanser 1997; Blahusch et al. 1999). In the division model, the undistorted point  $(x_u, y_u)^\top$  is computed from the distorted point as follows:

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \frac{1}{1 + \kappa r_d^2} \begin{pmatrix} x_d \\ y_d \end{pmatrix}, \tag{10}$$

where  $r_d^2 = x_d^2 + y_d^2$ . The division model has the advantage that it can be inverted analytically. The distorted point can be computed by:

$$\begin{pmatrix} x_d \\ y_d \end{pmatrix} = \frac{2}{1 + \sqrt{1 - 4\kappa r_u^2}} \begin{pmatrix} x_u \\ y_u \end{pmatrix}, \tag{11}$$

where  $r_u^2 = x_u^2 + y_u^2$ . Note that the division model only supports radial distortion.

The second model that is supported is the polynomial model proposed by Brown (1966, 1971), which models radial as well as decentering distortions. The undistorted point is computed by:

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \begin{pmatrix} x_d (1 + K_1 r_d^2 + K_2 r_d^4 + K_3 r_d^6) + (P_1 (r_d^2 + 2x_d^2) + 2P_2 x_d y_d) \\ y_d (1 + K_1 r_d^2 + K_2 r_d^4 + K_3 r_d^6) + (2P_1 x_d y_d + P_2 (r_d^2 + 2y_d^2)) \end{pmatrix}. \tag{12}$$

The polynomial model cannot be inverted analytically. The computation of the distorted point from the undistorted point must be performed by a numerical root finding algorithm. This is significantly slower than the analytical inversion in the division model.

In our experience, these two models are accurate enough to handle all lenses that are typically encountered in machine vision applications. However, it would be easy to extend the camera model with more general models that treat distortions as transformations in the image plane, e.g., the rational distortion model (Claus and Fitzgibbon 2005), should the need arise.

Finally, the distorted point  $(x_d, y_d)^\top$  is transformed into the image coordinate system:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \frac{x_d}{s_x} + c_x \\ \frac{y_d}{s_y} + c_y \end{pmatrix}. \tag{13}$$

Here,  $s_x$  and  $s_y$  denote the pixel sizes on the sensor (more accurately: the pixel pitch) and  $(c_x, c_y)^\top$  is the principal point. Note that  $x$  refers to the horizontal axis of the image (increasing rightward) and  $y$  to the vertical axis (increasing downward). If we use homogeneous coordinates, we can

represent the transformation in (13) by the  $3 \times 3$  calibration matrix

$$\mathbf{K} = \begin{pmatrix} 1/s_x & 0 & c_x \\ 0 & 1/s_y & c_y \\ 0 & 0 & 1 \end{pmatrix}. \quad (14)$$

The camera model therefore consists of the following parameters:

- The six parameters of the exterior orientation (modeling the pose of the calibration objects in the  $n_o$  images):  $\alpha_l, \beta_l, \gamma_l, t_{l,x}, t_{l,y}$ , and  $t_{l,z}$ .
- The six parameters of the relative orientation of the  $n_c$  cameras with respect to camera 1:  $\alpha_k, \beta_k, \gamma_k, t_{k,x}, t_{k,y}$ , and  $t_{k,z}$ .
- The interior orientation of each camera (for simplicity, we omit the subscripts  $k$  here):  $c$  or  $m$ ;  $\kappa$  or  $K_1, K_2, K_3, P_1, P_2$ ;  $s_x, s_y, c_x$ , and  $c_y$ .

The above parameterization is very intuitive for machine vision users. All parameters have a physical meaning that is easy to understand. Approximate initial values for the interior orientation parameters can simply be read off the data sheets of the camera ( $s_x, s_y$ ) and the lens ( $c$  or  $m$ ) or can be obtained easily otherwise (the initial values for the principal point can be set to the center of the image and the distortion coefficients can typically be set to 0). Furthermore, the calibration results are easy to check for validity.

## 6.2 Model Degeneracies

*Remark 1* The camera models above are overparameterized. We cannot determine  $c$  or  $m, s_x$ , and  $s_y$  uniquely since they functionally depend on each other. We can solve this problem by leaving  $s_y$  at the initial value the user specifies. This is a valid approach since camera signals are always transmitted line-synchronously, i.e., the acquired image necessarily has the same vertical pixel pitch as the image on the sensor. For  $s_x$ , this is not always the case. If the image is transmitted via an analog signal, the image may not be sampled pixel-synchronously by the frame grabber (Steger et al. 2008, Chapter 3.9).

Another method to remove the overparameterization that is frequently used is to move the parameters  $c$  and  $m$  from the projection equations (6) and (8) to the transformation into the pixel coordinates (13). Therefore, (6) becomes

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \frac{1}{z_k} \begin{pmatrix} x_k \\ y_k \end{pmatrix} \quad (15)$$

with projection matrix

$$\mathbf{P}_p = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (16)$$

(8) becomes

$$\begin{pmatrix} x_u \\ y_u \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} \quad (17)$$

with projection matrix

$$\mathbf{P}_o = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (18)$$

and (13) becomes

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} a_x x_d + c_x \\ a_y y_d + c_y \end{pmatrix} \quad (19)$$

with calibration matrix

$$\mathbf{K} = \begin{pmatrix} a_x & 0 & c_x \\ 0 & a_y & c_y \\ 0 & 0 & 1 \end{pmatrix}, \quad (20)$$

where,  $(a_x, a_y) = (c/s_x, c/s_y)$  for cameras that are perspective in object space and  $(a_x, a_y) = (m/s_x, m/s_y)$  for cameras that are telecentric in object space. We have used this kind of parameterization in the past for telecentric cameras (Steger et al. 2008, Chapter 3.9; Blahusch et al. 1999). However, feedback from users has shown us that this mingling of parameters is difficult to use for many users.

The interesting aspect of the above discussion is that we are free to use the parameters  $c$  and  $m$  in the projection matrices or the calibration matrices according to our needs. This enables us to separate all the parts of the model that are purely image-based transforms from everything that requires the poses and the projection. We will make use of this fact in several proofs below. Note that if we move  $c$  and  $m$  to the calibration matrix, we must scale the distortion coefficients appropriately. For example,  $\kappa$  and  $K_1$  must be divided by  $c^2$  or  $m^2$ ,  $K_2$  must be divided by  $c^4$  or  $m^4$ , etc.

*Remark 2* For lenses that are telecentric in object space, it is impossible to determine the principal point  $(c_x, c_y)^\top$  uniquely if there are no distortions. In other words, the principal point is solely determined by the distortions for this kind of lenses. If there are no distortions, the cameras are actually affine cameras in the sense of (Hartley and Zisserman 2003, Chapter 6.3). For these cameras,  $(c_x, c_y)$  and  $(t_{l,x}, t_{l,y})^\top$  have the same effect. As described in (Hartley and Zisserman 2003, Chapter 6.3.3), the appropriate solution is to leave



$(c_x, c_y)^\top$  at the initial values specified by the user (typically, the center of the image).

*Remark 3* As described in Sect. 9, we calibrate the cameras using a planar calibration object. For perspective lenses, it is well known that for some configurations of poses of the calibration object, not all parameters of the camera model can be determined simultaneously (Sturm and Maybank 1999). These degeneracies must be taken into account by the user when acquiring images of the calibration object.

*Remark 4* For telecentric lenses, the pose parameter  $t_{l,z}$  obviously cannot be determined. We arbitrarily set it to 1 m. In addition, in a multi-view setup, the relative pose parameters  $(t_{k,x}, t_{k,y}, t_{k,z})$  cannot be determined uniquely since all cameras can be moved arbitrarily along their optical axes without changing the image geometry. To provide a well-defined relative pose, we move the origins of the camera coordinate systems along the respective optical axes to a sphere with radius 1 m. The center of the sphere is given by a point that lies at a distance of 1 m on the optical axis in front of the reference camera.

*Remark 5* For telecentric cameras and planar calibration objects, the rotation part of the pose can only be determined up to a twofold ambiguity from a single camera. For example, a plane rotated by  $\alpha_l = 20^\circ$  looks identical to a plane rotated by  $\alpha_l = -20^\circ$ . This essentially is the same as a Necker reversal (Hartley and Zisserman 2003, Chapter 14.6). In a multi-view setup, these individual pose ambiguities can be resolved, albeit only up to an overall Necker reversal, which will have to be resolved manually by the user.

## 7 Models for Cameras with Tilt Lenses

### 7.1 Camera Models

We will now extend the camera models of Sect. 6 to handle tilt lenses correctly. To do so, we must model the transformation that occurs when the image plane is tilted for lenses that are telecentric in image space (image-side telecentric and bilateral telecentric lenses) and for lenses that are perspective in image space (perspective and object-side telecentric lenses).

The camera models in Sect. 6 have proven their ability to model standard cameras correctly for many years. Therefore, the tilt camera models should reduce to the standard models if the image plane is not tilted. From the discussion at the end of Sect. 4, we can see that the parameters  $c$  and  $m$  essentially model the ray angles in object space correctly.<sup>9</sup> Furthermore,

<sup>9</sup> For telecentric cameras, the rays are all parallel since they go through the center of the entrance pupil at infinity. Hence, they are all translated with respect to each other. A translation can also be regarded as a rotation

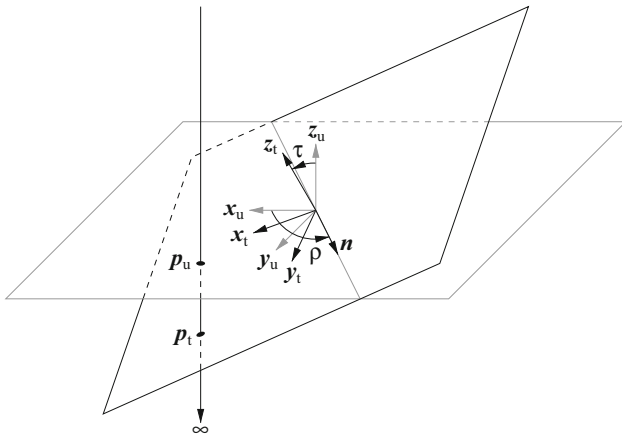
as discussed by Sturm et al. (2010, Page 43), the distortion models discussed above essentially model distortions of ray angles with respect to the optical axis. In the above distortion models, the rays in image space are represented by their intersections with a plane that is perpendicular to the optical axis. This is convenient for untilted image planes, since this plane is already available: it is the image plane. Since the optical axis of the lens is unaffected by a tilt of the image plane, we can still use the above mechanism to represent the distortions, which models the distortions of ray angles with respect to the optical axis by way of their intersections with a plane that is perpendicular to the optical axis. Since the actual image plane is now tilted, the untilted image plane becomes a virtual image plane that is used solely for the purpose of representing ray angles with respect to the optical axis and thus to compute the distortions. Consequently, these two parts of the model, i.e., the modeling of the ray angles in object space by  $c$  or  $m$  and the modeling of the distortions in a plane that is perpendicular to the optical axis, can remain unchanged.

Both tilt models therefore work by projecting a point from a virtual image plane that is perpendicular to the optical axis to the tilted image plane. Therefore, we first describe how we can model the pose of the tilted image plane in a manner that is easy to understand for the user. As discussed at the end of Sect. 3, almost all tilt lenses work by first selecting the direction in which to tilt the lens and then tilting the lens. The selection of the direction in which to tilt essentially determines a rotation axis  $\mathbf{n}$  in a plane orthogonal to the optical axis, i.e., the untilted image plane, and then rotating the image plane around that axis. Let the image coordinate system of the untilted image plane be given by the vectors  $\mathbf{x}_u$  and  $\mathbf{y}_u$ . We can extend this image coordinate system to a 3D coordinate system by the vector  $\mathbf{z}_u$ , which points back to the scene along the optical axis. Figure 14 displays the untilted image plane and this coordinate system in medium gray. The axis  $\mathbf{n}$  around which the image plane is tilted can be parameterized by the angle  $\rho$  ( $0 \leq \rho < 2\pi$ ) as follows:

$$\mathbf{n} = \begin{pmatrix} \cos \rho \\ \sin \rho \\ 0 \end{pmatrix}. \quad (21)$$

If we rotate the coordinate system  $(\mathbf{x}_u, \mathbf{y}_u, \mathbf{z}_u)$  by the tilt angle  $\tau$  ( $0 \leq \tau < \pi/2$ ) around  $\mathbf{n}$ , the coordinate axes  $(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$  of the tilted image plane are given in the coordinate system of the untilted image plane by (see Spong et al. 2006, Chapter 2.5.3)

around a point at infinity. Therefore, when we speak about ray angles, we will use this extended interpretation of the term angle to avoid tedious repetitions of the same case distinctions.



**Fig. 14** The projection of a point  $p_u$  from the untilted image plane to a point  $p_t$  in the tilted image plane for a camera that is telecentric in image space. The coordinate system of the untilted image plane is given by  $(x_u, y_u, z_u)$ , that of the tilted image plane by  $(x_t, y_t, z_t)$ . The camera's viewing direction is along the  $z_u$  axis, which points towards the scene. The rotation axis around which the image plane is tilted is given by  $n$ , which forms an angle  $\rho$  with  $x_u$ . The image plane is tilted by the angle  $\tau$  around the axis  $n$ . The direction of the light ray is indicated by the vertical arrow, which goes through the image-side projection center (the exit pupil), located at infinity. Note that the image plane is shown in its correct orientation in the camera: upside-down

$$\mathbf{R}_t = \begin{pmatrix} x_t & y_t & z_t \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} c_\rho^2(1 - c_\tau) + c_\tau & c_\rho s_\rho(1 - c_\tau) & s_\rho s_\tau \\ c_\rho s_\rho(1 - c_\tau) & s_\rho^2(1 - c_\tau) + c_\tau & -c_\rho s_\tau \\ -s_\rho s_\tau & c_\rho s_\tau & c_\tau \end{pmatrix}, \quad (22)$$

where  $c_\theta = \cos \theta$  and  $s_\theta = \sin \theta$ . Note that the rotation matrix  $\mathbf{R}_t$  also represents a transformation of points from the tilted coordinate system  $(x_t, y_t, z_t)$  to the untilted coordinate system  $(x_u, y_u, z_u)$ .

Note that the semantics of the tilt parameters are quite easy to understand. A rotation angle  $\rho = 0^\circ$  means that the lens (i.e., the optical axis) is tilted downwards by  $\tau$  with respect to the camera housing, for  $\rho = 90^\circ$ , it is tilted leftwards, for  $\rho = 180^\circ$  upwards, and for  $\rho = 270^\circ$  rightwards.

To project a point  $p_u = (x_u, y_u)^\top$  from the untilted image plane to a point  $p_t = (x_t, y_t)^\top$  in the tilted image plane, it is easiest to consider the inverse of this projection, i.e., to compute the orthographic projection from the tilted to the untilted image plane. It is well known that the orthographic projection of one vector onto another vector is given by the scalar product of the two vectors and that a linear transformation is given by projecting the axes of one coordinate system onto the axes of the other coordinate system (see, e.g., Spong et al. 2006, Chapter 2.2). Therefore, to compute the transformation from the tilted image plane to the untilted image plane, we must compute the projections (scalar products) of  $x_t$  and  $y_t$  onto  $x_u$  and  $y_u$ . For example, the element  $h_{11}$  of the projection matrix

is the projection of  $x_t$  onto  $x_u$ . Since  $x_u = (1, 0, 0)^\top$  and  $x_t = (r_{11}, r_{21}, r_{31})^\top$  in the coordinate system of the untilted image plane, the projection is given by  $x_t \cdot x_u = r_{11}$ . If we do this for all the  $x$  and  $y$  coordinate system vectors, we see that the inverse tilt transformation is an affine transformation:

$$\mathbf{H}_0^{-1} = \begin{pmatrix} r_{11} & r_{12} & 0 \\ r_{21} & r_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (23)$$

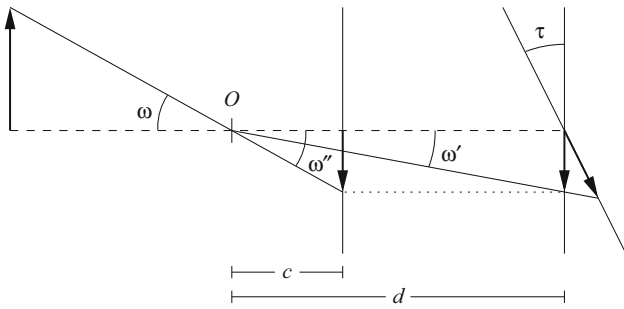
Therefore, the transformation from untilted to tilted image coordinates for lenses that are telecentric in image space is given by

$$\mathbf{H}_0 = \begin{pmatrix} \frac{r_{22}}{r_{11}r_{22} - r_{12}r_{21}} & -\frac{r_{12}}{r_{11}r_{22} - r_{12}r_{21}} & 0 \\ -\frac{r_{21}}{r_{11}r_{22} - r_{12}r_{21}} & \frac{r_{11}}{r_{11}r_{22} - r_{12}r_{21}} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (24)$$

We have  $p_t = \mathbf{H}_0 p_u$ . We insert this tilt transformation into the camera model between the distortion (10)–(12) and the transformation to the image coordinate system (13). Consequently, the world points are projected to the untilted image plane, distorted within the untilted image plane, transformed to the tilted image plane, and then transformed to the image coordinate system. Note that the tilt transformation is invariant to the choice of units. Consequently, the fact that we can move  $c$  or  $m$  to the transformation to the image plane, mentioned at the end of Sect. 6, is still valid.

We now turn to cameras with lenses that are perspective in image space. Here, we must be able to model the different ray angles in object and image space correctly. From the discussion in Sect. 4 and Fig. 13, it is evident what must be done to model the different ray angles correctly: we must locate the untilted image plane at the true distance from the center of the exit pupil. This distance was called  $d'$  in Fig. 13. For simplicity, we will call it  $d$  from now on. We will refer to  $d$  as the image plane distance. We require  $0 < d < \infty$ . Figure 15 displays this geometry. Points in object space are first projected to a virtual image plane that is orthogonal and lies at a distance  $c$  from the projection center  $O$ . Obviously, this causes the object and image space ray angles to be identical ( $\omega'' = \omega$ ) and therefore causes the object space ray angles  $\omega$  to be modeled correctly. To model the image space ray angles correctly, the virtual image plane is shifted to a distance  $d$  (which corresponds to  $d'$  in Fig. 13), resulting in the correct image space ray angles  $\omega' \neq \omega$ .<sup>10</sup> Note that this shift does not change the virtual image in any way. Next, the points are distorted in the virtual image plane. With the virtual image at its correct distance  $d$ , the plane can now be tilted by the correct tilt angle  $\tau$ .

<sup>10</sup> Geometrically, this can also be regarded as an orthogonal projection from the virtual image plane at  $c$  to that at  $d$ .



**Fig. 15** The ray geometry of the perspective tilt camera model. Points in object space are first projected to a virtual image plane that is orthogonal and lies at a distance  $c$  from the projection center  $O$ . This causes the object and image space ray angles to be identical ( $\omega'' = \omega$ ) and therefore causes the object space ray angles  $\omega$  to be modeled correctly. To model the image space ray angles correctly, the virtual image plane is shifted to a distance  $d$  (which corresponds to  $d'$  in Fig. 13), resulting in the correct image space ray angles  $\omega' \neq \omega$ . Next, the points are distorted in the virtual image plane. With the virtual image at its correct distance  $d$ , the plane can now be tilted by the correct tilt angle  $\tau$

With this, the major difference to the telecentric model, as shown in Fig. 16, is that the image-side projection center (the center of the exit pupil) lies at a finite distance  $d$  in front of the image plane.

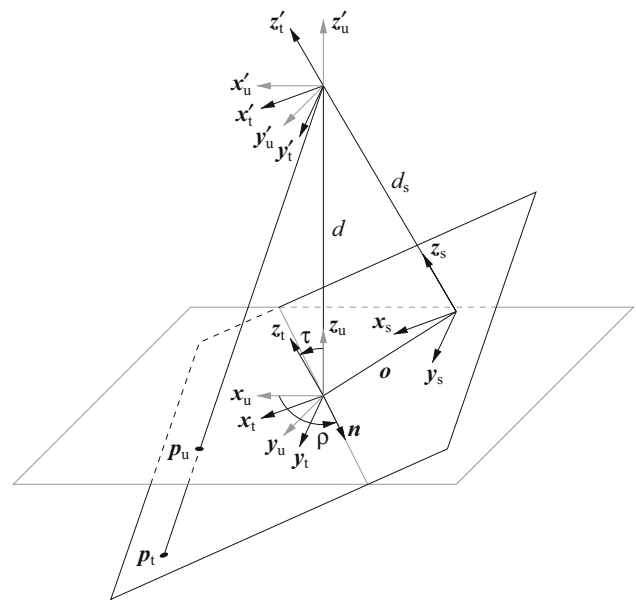
As shown in Fig. 16, the coordinate systems  $(x_u, y_u, z_u)$  and  $(x_t, y_t, z_t)$  have the same meaning as in the telecentric case. If we shift the two coordinate systems to the projection center and construct another coordinate system  $(x_s, y_s, z_s)$  in the tilted image plane that is parallel to  $(x_t, y_t, z_t)$  and lies at the perpendicular projection of the projection center into the tilted image plane, we see that  $(x_u, y_u, z_u)$  and  $(x_s, y_s, z_s)$  are equivalent to two perspective cameras that are rotated around their common projection center. Therefore, we can construct two calibration matrices  $\mathbf{K}_u$  and  $\mathbf{K}_s$  and a rotation matrix  $\mathbf{R}$  that relate the two cameras. Then, the projection from  $(x_u, y_u, z_u)$  to  $(x_s, y_s, z_s)$  is given by  $\mathbf{K}_s \mathbf{R} \mathbf{K}_u^{-1}$  (Hartley and Zisserman 2003, Chapter 8.4.2). Finally, we must shift  $(x_s, y_s, z_s)$  to  $(x_t, y_t, z_t)$  by a translation  $\mathbf{o}$  within the tilted image plane. This translation can be modeled by a translation matrix  $\mathbf{T}$ . Thus, the complete projection is given by

$$\mathbf{H}_p = \mathbf{T} \mathbf{K}_s \mathbf{R} \mathbf{K}_u^{-1}. \tag{25}$$

The calibration matrix  $\mathbf{K}_u$  is obviously given by

$$\mathbf{K}_u = \begin{pmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{26}$$

The distance  $d_s$  of  $(x_s, y_s, z_s)$  from the projection center is the orthogonal projection of the vector  $\mathbf{d} = (0, 0, d)^\top$  onto the axis  $z_s = z_t$ , i.e.,  $d_s = \mathbf{d} \cdot \mathbf{z}_t = dr_{33}$ , with the rotation matrix  $\mathbf{R}_t$  as defined in (22). Therefore, we have



**Fig. 16** The projection of a point  $p_u$  from the untilted image plane to a point  $p_t$  in the tilted image plane for a camera that is perspective in image space. The coordinate system of the untilted image plane is given by  $(x_u, y_u, z_u)$ , that of the tilted image plane by  $(x_t, y_t, z_t)$ . The camera’s viewing direction is along the  $z_u$  axis, which points towards the scene. The rotation axis around which the image plane is tilted is given by  $\mathbf{n}$ , which forms an angle  $\rho$  with  $x_u$ . The image plane is tilted by the angle  $\tau$  around the axis  $\mathbf{n}$ . These two coordinate systems can also be attached to the image-side projection center (the center of the exit pupil) as  $(x'_u, y'_u, z'_u)$  and  $(x'_t, y'_t, z'_t)$ . The distance from the projection center to the intersection of the optical axis with the untilted and tilted image planes is  $d$ . The coordinate system  $(x_s, y_s, z_s)$  lies in the tilted image plane at the perpendicular projection of the projection center. It lies at the distance  $d_s$  from the projection center, has the same orientation as  $(x_t, y_t, z_t)$ , and is offset from this coordinate system by the vector  $\mathbf{o}$ . Note that the image plane is shown in its correct orientation in the camera: upside-down

$$\mathbf{K}_s = \begin{pmatrix} dr_{33} & 0 & 0 \\ 0 & dr_{33} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{27}$$

The rotation matrix  $\mathbf{R}$  in (25) must transform points from  $(x_u, y_u, z_u)$  to points in  $(x_s, y_s, z_s)$ , i.e., in  $(x_t, y_t, z_t)$ . As discussed above, the matrix  $\mathbf{R}_t$  in (22) performs the inverse of this transformation. Thus,

$$\mathbf{R} = \mathbf{R}_t^\top. \tag{28}$$

Finally, the translation vector  $\mathbf{o}$  is the negative of the orthogonal projection of  $\mathbf{d}$  onto  $x_t$  and  $y_t$ , i.e.,  $\mathbf{o} = -(\mathbf{d} \cdot \mathbf{x}_t, \mathbf{d} \cdot \mathbf{y}_t)^\top = -(dr_{31}, dr_{32})^\top$ . Therefore,

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & -dr_{31} \\ 0 & 1 & -dr_{32} \\ 0 & 0 & 1 \end{pmatrix}. \tag{29}$$

By Substituting (26)–(29) into (25), we obtain

$$\mathbf{H}_p = \begin{pmatrix} r_{11}r_{33} - r_{13}r_{31} & r_{21}r_{33} - r_{23}r_{31} & 0 \\ r_{12}r_{33} - r_{13}r_{32} & r_{22}r_{33} - r_{23}r_{32} & 0 \\ r_{13}/d & r_{23}/d & r_{33} \end{pmatrix}. \quad (30)$$

As above, we insert this tilt transformation into the camera model between the distortion (10)–(12) and the transformation to the image coordinate system (13).

## 7.2 Model Properties and Degeneracies

**Proposition 1** *The tilt camera models reduce to the standard camera models if the image plane is not tilted.*

*Proof* Inserting  $\tau = 0$  into (22), we obtain  $\mathbf{R}_t = \mathbf{I}$ . Therefore,  $\mathbf{H}_o = \mathbf{I}$  and  $\mathbf{H}_p = \mathbf{I}$ .  $\square$

**Corollary 1** *For lenses that are perspective in image space, the image plane distance  $d$  cannot be determined if the image plane is not tilted.*

*Proof* This is obvious from Proposition 1 and the discussion in Sect. 4. This also means that the user should select the standard perspective camera model if there is no tilt.  $\square$

**Remark 6** A consequence of Corollary 1 is that the smaller  $\tau$  is, the less precisely  $d$  can be determined. If the calibration has converged with a small RMS error, this is of no concern since the calibrated system will be consistent with the imaging geometry.

**Proposition 2** *The tilt homographies in (24) and (30) are consistent with each other, i.e.,  $\mathbf{H}_p \rightarrow \mathbf{H}_o$  as  $d \rightarrow \infty$ .*

*Proof* By substituting (22) into (24) and (30) and simplifying the trigonometric terms, we obtain

$$\mathbf{H}_o = \begin{pmatrix} \frac{c_\rho^2 c_\tau + s_\rho^2}{c_\tau} & \frac{c_\rho s_\rho (c_\tau - 1)}{c_\tau} & 0 \\ \frac{c_\rho s_\rho (c_\tau - 1)}{c_\tau} & \frac{s_\rho^2 c_\tau + c_\rho^2}{c_\tau} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (31)$$

and

$$\mathbf{H}_p = \begin{pmatrix} c_\rho^2 c_\tau + s_\rho^2 & c_\rho s_\rho (c_\tau - 1) & 0 \\ c_\rho s_\rho (c_\tau - 1) & s_\rho^2 c_\tau + c_\rho^2 & 0 \\ s_\rho s_\tau / d & -c_\rho s_\tau / d & c_\tau \end{pmatrix}. \quad (32)$$

If we dehomogenize  $\mathbf{H}_p$  by dividing by  $c_\tau$ , the result follows.  $\square$

**Remark 7** Lenses with  $d \geq 1000$  m can be regarded as telecentric in image space for all practical purposes.

**Remark 8** It is trivial to show that in the perspective tilt camera model,  $c$  and  $d$  are related to the pupil magnification factor (2) by

$$m_p = \frac{d}{c}. \quad (33)$$

**Proposition 3** *If we denote the orthographic tilt homography in (31) as  $\mathbf{H}_o(\rho, \tau)$  to make the parameters explicit, we have*

$$\mathbf{H}_o(\rho, -\tau) = \mathbf{H}_o(\rho, \tau) \quad (34)$$

$$\mathbf{H}_o(\rho + \pi, \tau) = \mathbf{H}_o(\rho, \tau) \quad (35)$$

$$\mathbf{H}_o(\rho + \pi, -\tau) = \mathbf{H}_o(\rho, -\tau). \quad (36)$$

*Proof* This follows directly from (31).  $\square$

**Remark 9** Proposition 3 shows that we cannot determine the tilt of the image plane uniquely for lenses that are telecentric in image space, even if we require  $\tau > 0$ . This is obvious from the geometry of the orthographic tilt homography. Note that this is the image space analog of the degeneracy described in Remark 5. The ambiguity can be resolved if we require  $0 \leq \rho < \pi$ . In practice, however, this would confuse the users, who typically would be surprised to have  $\rho$  reduced to this range if they specified a value of  $\rho$  outside this interval as the initial value (e.g., if the calibration returned that the lens is tilted right when the initial values specified that it is tilted left). Since the calibration will typically converge to the same half space in which the initial  $\rho$  was specified, we do not reduce  $\rho$  modulo  $\pi$ .

**Proposition 4** *If we denote the perspective tilt homography in (32) as  $\mathbf{H}_p(\rho, \tau, d)$  to make the parameters explicit, we have*

$$\mathbf{H}_p(\rho, -\tau, d) = \mathbf{H}_p(\rho + \pi, \tau, d) \quad (37)$$

$$\mathbf{H}_p(\rho, \tau, -d) = \mathbf{H}_p(\rho, -\tau, d) \quad (38)$$

$$\mathbf{H}_p(\rho, \tau, -d) = \mathbf{H}_p(\rho + \pi, \tau, d). \quad (39)$$

*Proof* This follows directly from (32).  $\square$

**Remark 10** Proposition 4 shows that there are no degeneracies in the perspective tilt homography if we require  $0 \leq \tau < \pi/2$ ,  $0 \leq \rho < 2\pi$ , and  $d > 0$ .

**Proposition 5** *If a lens that is perspective in image space is tilted around the horizontal or vertical axis, i.e., if  $\rho \in \{0, \pi/2, \pi, 3\pi/2\}$ , the values of  $\tau$ ,  $d$ ,  $s_x$ , and  $s_y$  cannot be determined uniquely. This degeneracy can be resolved if the pixel aspect ratio is known.*

*Proof* We will only prove the case  $\rho = 0$ . The other cases can be proved in an analogous manner. The calibration matrix  $\mathbf{K}$



in (14) can be split into a scaling and a translation part. Only the scaling part is relevant for this proof, i.e., we can set  $c_x = c_y = 0$ . If we multiply  $\mathbf{K}(s_x, s_y)$  and  $\mathbf{H}_p(0, \tau, d)$ , we obtain

$$\mathbf{K}(s_x, s_y)\mathbf{H}_p(0, \tau, d) = \begin{pmatrix} \cos \tau/s_x & 0 & 0 \\ 0 & 1/s_y & 0 \\ 0 & -\sin \tau/d & \cos \tau \end{pmatrix}. \quad (40)$$

Suppose one set of parameters  $(s_{x,1}, s_{y,1}, \tau_1, d_1)$  is given. We want to find a different solution  $(s_{x,2}, s_{y,2}, \tau_2, d_2)$ . Since (40) represents a homogeneous matrix, we require

$$\mathbf{K}(s_{x,1}, s_{y,1})\mathbf{H}_p(0, \tau_1, d_1) = s\mathbf{K}(s_{x,2}, s_{y,2})\mathbf{H}_p(0, \tau_2, d_2), \quad (41)$$

where  $s$  is a scaling factor. This results in four equations:

$$\cos \tau_1/s_{x,1} = s \cos \tau_2/s_{x,2} \quad (42)$$

$$1/s_{y,1} = s/s_{y,2} \quad (43)$$

$$\sin \tau_1/d_1 = s \sin \tau_2/d_2 \quad (44)$$

$$\cos \tau_1 = s \cos \tau_2. \quad (45)$$

We can select one of the four parameters on the right hand side of (41) arbitrarily, e.g.,  $\tau_2$ . We can then solve (45) for  $s$ :

$$s = \frac{\cos \tau_1}{\cos \tau_2}. \quad (46)$$

Substituting  $\cos \tau_1$  from (45) into (42) and solving for  $s_{x,2}$  yields

$$s_{x,2} = s_{x,1}. \quad (47)$$

Substituting  $s$  from (46) into (43) and solving for  $s_{y,2}$  yields

$$s_{y,2} = s_{y,1} \frac{\cos \tau_1}{\cos \tau_2}. \quad (48)$$

Finally, substituting  $s$  into (44) and solving for  $d_2$  yields

$$d_2 = d_1 \frac{\tan \tau_2}{\tan \tau_1}. \quad (49)$$

Therefore, we can find a valid solution for any chosen  $\tau_2$  by (47)–(49). From (48), we see that a unique solution can be enforced by requiring  $s_{y,2} = s_{y,1}$ , i.e., by assuming that the aspect ratio of the pixels is known.  $\square$

*Remark 11* We have shown that Proposition 5 holds under the assumption that  $s_x$  and  $s_y$  can vary. As discussed in Remark 1, this implies that the degeneracy is present even if  $s_y$  is fixed since the effect of  $s_x$  and  $s_y$  can also be modeled by  $c$  or  $m$  and  $s_x$ .

*Remark 12* Instead of selecting  $\tau_2$ , we could also have selected  $d_2$  and solved for the remaining parameters. In particular, we would then have obtained

$$\tau_2 = \arctan\left(\frac{d_2}{d_1} \tan \tau_1\right). \quad (50)$$

If we regard  $d_1 = d$  as the true image plane distance and imagine that we force the image space ray angles to be identical to the object space ray angles by setting  $d_2 = c$ , we can see that the tilt angle  $\tau_2$  will be incorrect if  $d \neq c$ . For example, if  $d$  is actually three times as large as  $c$  and if  $\tau_1 = 5^\circ$ , we have  $\tau_2 \approx 1.67^\circ$ . From (48), it follows that  $s_{y,2}$  will be wrong by about  $-0.34\%$ .

*Remark 13* Proposition 5 and Remark 12 show that for all the perspective tilt camera models that force the ray angles in object and image space to be identical (see Sect. 5.1), the following properties hold if the ray angles are actually different:

- If the tilt is in the vertical or horizontal direction and the approach calibrates the pixel aspect ratio, the calibration will converge with a low RMS error, i.e., the model will be consistent with the imaging geometry, but the tilt angle and the aspect ratio of the pixels will be wrong.
- If the tilt is in the vertical or horizontal direction and the approach does not calibrate the pixel aspect ratio, the calibration will converge with a large RMS error, i.e., the model will be inconsistent with the imaging geometry, and at least the tilt angle will be wrong.
- If the tilt is not in the vertical or horizontal direction, the calibration will converge with a large RMS error, i.e., the model will be inconsistent with the imaging geometry, and at least the tilt angle will be wrong.

*Remark 14* Proposition 5 shows that the aspect ratio must be fixed in the calibration if the tilt is in the horizontal or vertical direction. This means that the user must exclude  $s_y$  from being calibrated if the tilt is close to the horizontal or vertical direction.

*Remark 15* The degeneracy in Proposition 5 is the image space analog to the fact that a regular camera cannot be calibrated from a single image of a planar calibration object that is rotated around the horizontal or vertical image axis, even if the principal point is known, as shown by Sturm and Maybank (1999).

**Proposition 6** For a lens that is telecentric in image space, the parameters  $c$  or  $m$ ,  $\rho$ ,  $\tau$ ,  $s_x$ , and  $s_y$  cannot be determined uniquely (even if  $s_y$  is fixed). This degeneracy can be resolved if  $s_x$  is known.

*Proof* For tilts in the horizontal or vertical direction, it is obvious that  $\mathbf{H}_o$  simply scales the image by a factor of  $1/\cos \tau$  in the horizontal or vertical direction. Obviously, this can also be modeled by  $c$  or  $m$  and  $s_x$ . Fixing  $s_x$  removes this degeneracy. For tilts in any other direction, the claim follows from Theorems 1 and 2, to be discussed in Sect. 8.  $\square$

*Remark 16* While the parameterization in (22) is very easy to understand for users, it has the disadvantage that  $\rho$  is undetermined if  $\tau = 0$ . We could simply set  $\rho = 0$  in this case, but this does not work in the camera calibration, where both parameters must be optimized. Therefore, internally we use the Rodrigues parameterization (Lenz 1988, Chapter 10.2; Morawiec and Field 1996), also called Gibbs or Cayley parameterization (Bauchau and Trainelli 2003). Here, the rotation is parameterized by  $\mathbf{r} = (r_x, r_y, r_z)^\top = \mathbf{n} \tan(\tau/2)$ . In our case,  $r_z = 0$  and we have:

$$\mathbf{R}_t = \frac{1}{1 + r_x^2 + r_y^2} \times \begin{pmatrix} 1 + r_x^2 - r_y^2 & 2r_x r_y & 2r_y \\ 2r_x r_y & 1 - r_x^2 + r_y^2 & -2r_x \\ -2r_y & 2r_x & 1 - r_x^2 - r_y^2 \end{pmatrix}. \quad (51)$$

The only singularity of the Rodrigues parameterization occurs for  $\tau = \pm\pi$ , a case that is of no interest to us. We convert the initial values of  $\tau$  and  $\rho$  specified by the user to  $\mathbf{r}$  for the calibration and convert the calibrated  $\mathbf{r}$  back to  $\tau$  and  $\rho$  on output. The price we pay to avoid the singularity is that the user no longer can force the tilt to occur in a particular direction since the tilt direction and tilt angle are jointly modeled by  $\mathbf{r}$  and cannot be excluded separately from the optimization.

*Remark 17* In the camera models we have proposed, the principal point is the intersection of the optical axis with the image plane. This point is sometimes called the auto-collimation point in photogrammetry (Luhmann et al. 2014, Chapter 3.3.2.2).

### 8 Tilt Cameras in Projective Geometry

As discussed in Sects. 6 and 7, the proposed models for cameras with tilt lenses can be represented by matrices if there are no distortions. In this section, we are only interested in a single camera with a single pose. Therefore, we will denote the pose matrix by  $\mathbf{H}_w$ . Thus, a perspective tilt lens camera is given by the camera matrix

$$\mathbf{P}_{pp} = \mathbf{K}\mathbf{H}_p\mathbf{P}_p\mathbf{H}_w, \quad (52)$$

where  $\mathbf{K}$ ,  $\mathbf{H}_p$ ,  $\mathbf{P}_p$ , and  $\mathbf{H}_w$  are given by (20), (32), (16), and (4), respectively. An image-side telecentric tilt lens camera

is given by

$$\mathbf{P}_{po} = \mathbf{K}\mathbf{H}_o\mathbf{P}_p\mathbf{H}_w, \quad (53)$$

where  $\mathbf{H}_o$  is given by (31). An object-side telecentric tilt lens camera is given by

$$\mathbf{P}_{op} = \mathbf{K}\mathbf{H}_p\mathbf{P}_o\mathbf{H}_w, \quad (54)$$

where  $\mathbf{P}_o$  is given by (18). Finally, a bilateral telecentric tilt lens camera is given by

$$\mathbf{P}_{oo} = \mathbf{K}\mathbf{H}_o\mathbf{P}_o\mathbf{H}_w. \quad (55)$$

In this section, we will examine how tilt cameras are related to projective cameras, affine cameras, and general cameras at infinity (see Hartley and Zisserman 2003, Chapters 6.1–6.3). In particular, we examine the question whether these three camera types can be interpreted as cameras with tilt lenses. The proofs of the theorems we will present make use of the dual image of the absolute conic (DIAC) (Hartley and Zisserman 2003, Chapter 8.5), given by

$$\omega^* = \mathbf{P}\mathbf{Q}_\infty^*\mathbf{P}^\top, \quad (56)$$

where  $\mathbf{Q}_\infty^* = \text{diag}(1, 1, 1, 0)$  is the canonical form of the absolute dual quadric (Hartley and Zisserman 2003, Chapter 3.7). (The function `diag` constructs a diagonal matrix with the specified elements.) This will allow us to remove the exterior orientation from the equations to be solved. If we denote the entries of  $\mathbf{P}$  by  $p_{ij}$ , the elements  $\omega_{ij}^*$  of the DIAC  $\omega^*$  are given by

$$\omega_{ij}^* = \sum_{k=1}^3 p_{ik}p_{jk} \quad (57)$$

Note that  $\omega^*$  is a symmetric matrix.

**Theorem 1** *Every affine camera can be regarded as a bilateral telecentric tilt camera with square pixels and principal point at  $(0, 0)^\top$ .*

*Proof* See Appendix A.1.  $\square$

*Remark 18* This proves Proposition 6 for bilateral telecentric tilt cameras, even if the pixels are non-square and if there are distortions. Let us assume that the pixels are non-square and that the aspect ratio  $b$  of the pixels is known and fixed. This can be modeled by a calibration matrix of the form  $\text{diag}(ba, a, 1)$ . We can multiply the camera matrix  $\mathbf{M}$  by  $\text{diag}(1/b, 1, 1)$  from the left to obtain a camera matrix with square pixels and then apply the approach given in the proof of Theorem 1. If there are distortions, Proposition 6 is true since the distortions are merely a transformation within

the untilted image plane. Consequently, they do not alter the essential projection properties of the camera, i.e., for bilateral telecentric tilt lenses, the locations of the entrance and exit pupil at infinity are not altered by the distortions. Therefore, the exterior orientation and the pose of the image plane are invariant to distortions.

**Theorem 2** *Every finite projective camera can be regarded as an image-side telecentric tilt camera with square pixels.*

*Proof* See Appendix A.2.  $\square$

*Remark 19* This proves Proposition 6 for image-side telecentric tilt cameras, even if the pixels are non-square and if there are distortions. The argument is almost identical to Remark 18. The only modification is that we must multiply  $\mathbf{K}_c \mathbf{M}$  from the left by the matrix  $\text{diag}(1/b, 1, 1)$ , where  $\mathbf{K}_c$  is given by (103).

**Theorem 3** *Every general camera at infinity can be regarded as an object-side telecentric tilt camera.*

*Proof* See Appendix A.3.  $\square$

*Remark 20* The space of general cameras at infinity has ten degrees of freedom: twelve for the entries of the camera matrix less one for the arbitrary scale of the matrix less one for the constraint that the left  $3 \times 3$  submatrix is singular. To remove the scale factor, one can also normalize the camera matrix by  $\|\mathbf{M}\|_F = 1$  using the Frobenius norm. Therefore, the space of general cameras at infinity can also be regarded as a subspace (given by  $\det \mathbf{M}_3 = 0$ ) of the unit sphere  $S^{11}$  in  $\mathbb{R}^{12}$  with antipodes identified (the latter is isomorphic to the projective space  $\mathbb{P}^{11}(\mathbb{R})$ ).

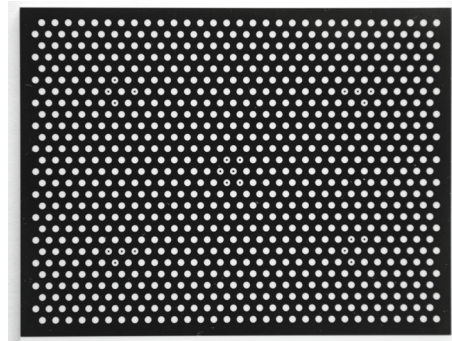
Theorem 3 establishes a minimal parameterization of the non-degenerate part of this space in terms of the parameters  $a_x, a_y, \tau, \rho, d, t_x, t_y$ , and three rotation parameters using a set of maps that are parameterized by a finite set of values of  $(c_x, c_y)^\top$ . This shows that the space has dimension 10, except in its degenerate part.

**Corollary 2** *Every projective camera can be regarded as a tilt camera.*

*Proof* This follows from Theorems 1–3. Affine cameras can be regarded as bilateral telecentric tilt cameras, general cameras at infinity can be regarded as object-side telecentric tilt cameras, and finite projective cameras can be regarded as image-side telecentric tilt cameras.  $\square$

## 9 Calibration

The camera calibration is performed using a planar calibration object. The advantages of planar calibration objects are



**Fig. 17** An example of the planar calibration object

that they are easier to manufacture accurately and that they can be handled more easily by the users. Furthermore, with planar calibration objects, it is very easy to define a plane in the world by simply placing the calibration object onto the plane of interest in the world. With this, it is possible to measure in world coordinates within the plane with a single camera, as described by Steger et al. (2008, Chapter 3.9.4). Another distinctive advantage of planar calibration objects is that they can be used conveniently in backlight applications if the calibration object is opaque and the control points are transparent (see, e.g., Steger et al. 2008, Chapter 4.7).

The calibration object uses a hexagonal layout of control points, as shown in Fig. 17. The hexagonal layout provides the highest density of control points of any layout. The calibration object has five finder patterns, indicated by the small dark circles within the control points, that facilitate the unique determination of the pose of the calibration object even if it is only partially visible in the image. This is advantageous since the calibration object can be imaged such that it covers the entire field of view of the camera. This increases the accuracy with which the camera parameters, in particular, the distortion parameters, can be determined.

Let the known 3D coordinates of the centers of the control points of the calibration object be denoted by  $\mathbf{p}_j$  ( $j = 1, \dots, n_m$ ).<sup>11</sup> As described in Sect. 6, the user conceptually acquires  $n_o$  images of the calibration object with each of the  $n_c$  cameras. The calibration object, however, does not need to be visible in all images simultaneously. If the calibration object is outside the field of view of a particular camera for a particular pose, no image needs to be acquired. There simply will be no observation of the calibration object in this case, which will be modeled appropriately in the optimization. However, there must be a chain of observations of the calibration object in multiple cameras that connects all the cameras. As described in Sect. 6, the pose (exterior orientation) of the calibration object in the reference camera is

<sup>11</sup> The calibration algorithm actually supports more than one calibration object. However, since this would make the notation even more complex, we will only discuss the case of a single calibration object.

denoted by  $\mathbf{H}_l$  ( $l = 1, \dots, n_o$ ) and the relative orientation of the cameras with respect to the reference camera is denoted by  $\mathbf{H}_k$  ( $k = 1, \dots, n_c$ ). Let us denote the corresponding parameters by the vectors  $\mathbf{e}_l$  and  $\mathbf{r}_k$ . Let the interior orientation of camera  $k$  be denoted by the vector  $\mathbf{i}_k$ . Furthermore, let the projection from world coordinates to the undistorted image plane, i.e., (3), (5), and, depending on the camera model, (6) or (8), be denoted by  $\pi_u$ . Let the projection from the image coordinate system back to the undistorted image plane, i.e., the inverse of (13), optionally the inverse of (24) or (30), depending on the camera model, and (10) or (12), depending on the camera model, be denoted by  $\pi_i^{-1}$ . In addition, let  $v_{jkl}$  denote a function that is 1 if the control point  $j$  of the observation  $l$  of the calibration object is visible with camera  $k$ , and 0 otherwise. The user effectively specifies the  $kl$  part of this function (albeit in a user-friendly manner) during the acquisition of the calibration images. The  $j$  part of this function is automatically determined during the extraction of the images of the control points, denoted by  $\mathbf{p}_{jkl}$ , from the calibration images.

To calibrate an arbitrary combination of cameras, the following function is minimized:

$$\varepsilon^2 = \sum_{l=1}^{n_o} \sum_{k=1}^{n_c} \sum_{j=1}^{n_m} v_{jkl} \|\pi_i^{-1}(\mathbf{p}_{jkl}, \mathbf{i}_k) - \pi_u(\mathbf{p}_j, \mathbf{e}_l, \mathbf{r}_k, \mathbf{i}_k)\|_2^2. \quad (58)$$

The minimization is performed by a suitable version of the sparse Levenberg–Marquardt algorithms described in (Hartley and Zisserman 2003, Appendix A6).<sup>12</sup>

The points  $\mathbf{p}_{jkl}$  are extracted by fitting ellipses (Fitzgibbon et al. 1999) to edges extracted with a subpixel-accurate edge extractor (Steger 1998, Chapter 3.3; Steger 2000). As discussed in Sect. 5.2, this causes a bias in the point positions. Section 10 will explain how this bias can be removed.

The Levenberg–Marquardt algorithm requires initial values for the camera parameters. One advantage of the camera model proposed in Sects. 6 and 7 is that the initial values for the interior orientation parameters can be directly obtained from the specifications of the camera and lens. For example, the initial values of  $s_x$  and  $s_y$  can be read off from the data sheet of the camera since they are the pixel size of the sensor elements. The initial value of  $c$  or  $m$  can be obtained

from the lens barrel or from the data sheet of the lens.<sup>13</sup> Some lens manufacturers provide Gaussian optics data for their lenses. If the lens manufacturer provides the distance of the exit pupil from the image plane, that value can be used as the initial value for  $d$ . If not, for perspective lenses the user can roughly estimate the pupil magnification factor  $m_p$  by eye by looking at the lens from the front and back and then use  $d = m_p c$  as the initial value for perspective lenses. For object-side telecentric lenses, the initial value for  $d$  typically corresponds roughly to the flange focal distance (for example, in the order of 20–50 mm for a C mount camera, for which the flange focal distance is 17.526 mm). Initial values for  $\tau$  and  $\rho$  are typically known from the considerations that led to the use of the tilt lens in the first place, i.e., from the Scheimpflug principle and Eq. (1). The initial value of  $(c_x, c_y)^T$  can be set to the center of the image and the initial values for the distortion parameters can be set to 0. In our experience, the choice of the initial parameters is not particularly critical as long as the initial value is within a relatively large factor, typically more than 5, of the actual value. Therefore, for example, the initial value of  $d$  for perspective tilt lenses can usually be set to the initial value of  $c$ .

One advantage of using circular marks is that initial parameters for the exterior orientation of the calibration object in each camera can be directly inferred from the shapes of the ellipses in the images (Lanser 1997, Chapter 3.2.1.2; Lanser et al. 1995). Initial values for the relative orientations of the cameras can be computed based on the initial values for the exterior orientations.

The calibration algorithm allows to exclude any parameter of the interior, exterior, and relative orientation globally or individually for each camera. The sole exception is that the tilt angles  $\rho$  and  $\tau$  can only be excluded jointly. This is a result of the implementation choice that internally the Rodrigues parameterization is used to represent the tilt angles (see Remark 16). The parameter exclusion mechanism is used to handle all the generic degeneracies described in Sects. 6.2 and 7.2 automatically. For example,  $s_x$  is excluded automatically if the user selects a bilateral or image-side telecentric tilt lens (see Proposition 6). Degeneracies that occur only for certain imaging geometries (e.g., the degeneracy described in Proposition 5 and Remarks 11, 12, and 14) must be handled by the user.

The proposed calibration algorithm provides some features that are novel and useful for the users. First of all, an arbitrary combination of perspective and telecentric cameras can be calibrated simultaneously, including their relative orientation. This facilitates a multi-view stereo reconstruction using an arbitrary combination of cameras. We will discuss this aspect in more detail in Sect. 12. Second, the formulation

<sup>12</sup> Like the vast majority of publications that describe the Levenberg–Marquardt algorithm, Hartley and Zisserman (2003, Appendix A6) only describe the minimization of residuals of the form  $\|\mathbf{y} - f(\mathbf{x})\|_2^2$ . However, the Levenberg–Marquardt algorithm can be used to minimize general squared residuals, and in particular residuals of the form  $\|g(\mathbf{x}, \mathbf{y}) - f(\mathbf{x})\|_2^2$  that is used in (58) (Nocedal and Wright 2006, Chapter 10).

<sup>13</sup> The focal length or magnification is typically printed onto the lens barrel by the manufacturers.



in (58) of minimizing the error in the undistorted image plane allows an arbitrary mixture of distortion models to be calibrated efficiently. In contrast, the typical projection to the image coordinate system would mean that the polynomial distortion model would have to be evaluated by a numerical root finding algorithm, which is significantly less efficient. Finally, the fact that the layout of the calibration object was designed so that it can fill the entire field of view and to maximize the density of the control points means that fewer calibration images must be acquired by the user to achieve the desired accuracy.

## 10 Bias Removal

In this section, we will describe two approaches to remove the control point bias.

The first approach is an extension of the approach by Heikkilä (2000). As discussed in Sect. 5.2, Heikkilä's approach does not consider distortion bias, which means that it does not fully remove the bias. Therefore, we extend his approach to correctly handle the distortion bias. The core idea to achieve this was already described by Lanser (1997, Chapter 3.2.1).<sup>14</sup> To remove the distortion bias, we transform the extracted subpixel-accurate control point contours to the undistorted image plane with the transformation  $\pi_i^{-1}$  (see Sect. 9) using the current set of interior orientation parameters. Ellipses are then fitted to the undistorted control point contours. Afterwards, the bias correction computed by the algorithm of Heikkilä (2000) is applied to the results of the ellipse fitting.<sup>15</sup> The complete algorithm first calibrates the cameras using the biased control point positions, applies the bias correction, and calibrates the cameras anew using the unbiased control point positions. The second step could be iterated. However, this typically results in insignificant improvements of the camera parameters, and therefore is not done in our implementation.

The above algorithm provides very accurate results, as the experiments below show. However, it does not work for line-scan cameras since for these cameras there is no undistorted image plane to which the contours could be rectified.<sup>16</sup> The second approach we propose can also be used for line-scan cameras. It basically is an efficient version of the approaches proposed by Datta et al. (2009) and Vo et al. (2011). To remove the bias, we transform the control point contours back to the plane of the calibration object with the approach

described by Steger et al. (2008, Chapter 3.9.4) using the current estimates of the interior orientation of the camera and the exterior orientation of the calibration object. Then, circles are fitted to the rectified contours. Finally, the centers of the fitted circles are projected to the undistorted image plane or the image coordinate system, as required. The camera is first calibrated using the biased control point positions, the bias is removed, and the camera is calibrated anew. The results of this approach are equally accurate as those of the first approach. For historical reasons, we use the first approach for area-scan cameras and the second approach for line-scan cameras.

To examine the potential accuracy improvement that can be achieved by the bias removal, we constructed a synthetic example in which we tried to maximize the perspective and distortion bias. We simulated a 20 megapixel camera with the same sensor specifications as a Canon EOS 6D and a very wide-angle 17 mm tilt/shift lens. To maximize the overall bias, we simulated a very large pincushion distortion. To maximize the perspective bias, 16 calibration images in which the calibration object was tilted by up to 45° and covered a large part of the field of view were simulated with an algorithm that assumes a perfect camera (no noise, 100 % fill factor, linear gray value response; see Steger et al. 2008, Chapter 3.7.4). This resulted in a diameter of the control points between 60 and 330 pixels in the calibration images, a perspective bias of up to 2.5 pixels (standard deviation (SD)  $\approx$  0.2 pixels), a distortion bias of up to 4.5 pixels (SD  $\approx$  0.4 pixels), and an overall bias of up to 6 pixels (SD  $\approx$  0.5 pixels). As can be seen from the standard deviations, despite the effort to maximize the bias, for most control points the bias is less than 1 pixel.

Since wide-angle lenses typically exhibit a barrel distortion, we also simulated this kind of distortion. This resulted in a perspective bias of up to 1.5 pixels (SD  $\approx$  0.15 pixels),<sup>17</sup> a distortion bias of up to 1.1 pixels (SD  $\approx$  0.1 pixels), and an overall bias of up to 0.7 pixels (SD  $\approx$  0.1 pixels). As can be seen, the distortion bias of barrel distortions cancels the perspective bias to some extent. Therefore, it is important to note that barrel distortions do not simulate the worst case.

The ground truth camera parameters and the results of the calibration without and with bias removal are displayed in Table 1. In addition to the interior orientation parameters, Table 1 also lists the maximum translation and angle errors of the exterior orientations of the calibration object. They were calculated by inverting the calibrated exterior orientations and composing them with the true exterior orientations. This

<sup>14</sup> Lanser, however, did not realize that his idea can be used to remove the distortion bias iteratively.

<sup>15</sup> This is only done for lenses that are perspective in object space. For lenses that are telecentric in object space, there is no perspective bias.

<sup>16</sup> The line-scan camera model we use is described in Steger et al. (2008, Chapter 3.9.2).

<sup>17</sup> The perspective bias is smaller for barrel distortions than for pincushion distortions since we distort the projected points and fitted ellipse centers by the radial distortions to measure them in the same space as the distortion bias, which we necessarily must measure after the radial distortions have been applied.

**Table 1** Results of the bias removal

	Ground truth	Biased results	Unbiased results
RMS error (Pixel)		0.074351	0.002860
$c$ (m)	0.024	0.023983795	0.024000059
$\kappa$ ( $\text{m}^{-2}$ )	500	500.5791428	499.9993895
$d$ (m)	0.05	0.049999964	0.050000069
$\tau$ ( $^\circ$ )	15	14.99723759	15.00000603
$\rho$ ( $^\circ$ )	30	30.02069524	29.99994620
$s_x$ ( $\mu\text{m Pixel}^{-1}$ )	6.55	6.55015328	6.54999932
$s_y$ ( $\mu\text{m Pixel}^{-1}$ )	6.55	6.55000000	6.55000000
$c_x$ (Pixel)	2636	2635.928262	2636.000462
$c_y$ (Pixel)	1874	1873.942690	1874.000215
$\varepsilon_t$ (mm)		0.1871209	0.0005827
$\varepsilon_a$ ( $^\circ$ )		0.0084054	0.0000463
RMS error (Pixel)		0.021138	0.003161
$c$ (m)	0.024	0.024000152	0.023999991
$\kappa$ ( $\text{m}^{-2}$ )	-500	-499.8946020	-500.0000536
$d$ (m)	0.05	0.049997673	0.049999958
$\tau$ ( $^\circ$ )	15	14.99988942	14.99999604
$\rho$ ( $^\circ$ )	30	30.00203993	30.00003883
$s_x$ ( $\mu\text{m Pixel}^{-1}$ )	6.55	6.55002037	6.55000049
$s_y$ ( $\mu\text{m Pixel}^{-1}$ )	6.55	6.55000000	6.55000000
$c_x$ (Pixel)	2636	2635.988750	2635.999484
$c_y$ (Pixel)	1874	1873.950543	1874.000014
$\varepsilon_t$ (mm)		0.0328418	0.0001116
$\varepsilon_a$ ( $^\circ$ )		0.0072942	0.0000341

The maximum translation error of the calibrated exterior orientations of the calibration object is given by  $\varepsilon_t$  and the maximum angle error by  $\varepsilon_a$ . The table shows the results of two experiments: a camera with a pin-cushion distortion ( $\kappa = 500$ ) and a camera with barrel distortion ( $\kappa = -500$ )

allows us to measure the errors in the coordinate system of the calibration object. The translation error is the length of the translation of the combined pose. To compute the angle error, the rotation of the combined pose is converted into an angle/axis representation and the absolute value of the angle is used as the angle error.

As can be seen from Table 1, the bias removal reduces the relative errors by 2–4 orders of magnitude in this example. A second iteration of the bias removal would only reduce the errors further by a factor of 2–3 and therefore is not worthwhile. We also note that the errors in the biased results are 1–2 orders of magnitude smaller if a barrel distortion of the same magnitude is used, while the errors in the unbiased results are up to one order of magnitude smaller.

The proposed approach to remove the bias is computationally very efficient. The calibration without bias removal requires 30 ms (excluding the time necessary to extract the control points in the calibration images). Including the bias removal increases the runtime by 700 ms. The transformation of the control point contours requires approximately 11 ms per image. In contrast, the approaches by [Datta et al.](#)

(2009) and [Vo et al.](#) (2011) would necessitate transforming the calibration images, which would require approximately 650 ms per image in our software. In addition, the control point extraction would have to be executed again for the rectified images. This would require approximately 300 ms per image in our software. Therefore, the proposed approach is more than 20 times more efficient than the approaches by [Datta et al.](#) (2009) and [Vo et al.](#) (2011).

## 11 Experiments

The proposed camera models have been evaluated extensively on numerous lenses. Standard SLR lenses were attached to a Novoflex Balpro T/S bellows, equipped with Nikon F-mount lens and camera adapters. This setup allows us to turn any Nikon SLR lens into a tilt/shift lens. However, since the minimum extension of the bellows that facilitates a tilt of approximately  $5^\circ$  is about 3 cm, the bellows acts like an extension tube. Thus, for lenses with a short focal length, the minimum object distance becomes impractically

small, which means that this setup only works for lenses with a sufficiently large focal length (larger than about 70 mm). Furthermore, we tested many commercially available SLR tilt/shift lenses. In addition, we constructed a special camera housing with a bellows that enabled us to attach most C-mount machine vision lenses to the housing and to be able to tilt the lenses around the vertical axis. In this section, we will present the most salient evaluation results.

Figure 18 shows a Sigma 150 mm F2.8 macro lens and a Nikon D3 camera attached to the Novoflex Balpro T/S bellows. The focus was set to a distance close to the minimum object distance of the lens. The lens has a focusing mechanism where the entrance pupil recedes within the lens as the focus is adjusted from infinity to the minimum object distance, whereas the exit pupil's size grows and its position remains approximately unchanged. This suggests that the lens has a principal distance that differs significantly between the infinity and near focus settings. At the chosen focus setting, the entrance pupil appears significantly larger than the exit pupil. Table 2 displays the calibration results using this setup.

To obtain a baseline for the RMS error, an experiment with no tilt ( $\tau \approx 0$ ) was performed. With the proposed camera model, this resulted in an RMS error of 0.12552. As described in Remark 6,  $d$  cannot be determined reliably if  $\tau \approx 0$ . This is why  $d$  has converged to an essentially arbitrary value in this case. Table 2 shows that, with the proposed perspective tilt camera model, the RMS error only increases slightly if the camera is tilted by an angle of  $\tau \approx 6^\circ$  around the axes  $\rho \approx 90^\circ$  and  $\rho \approx 135^\circ$ . The calibration results reflect that the entrance and exit pupils have different sizes. The image plane distance  $d$  is significantly smaller than the principal distance  $c$ . By (33), the pupil magnification factor of this lens at this particular focus setting is  $\approx 0.3$ . Note that the principal distance  $c$  is much larger than the nominal 150 mm because of the focusing mechanism of the lens and because the bellows acts as an extension tube.

The middle and lower parts of Table 2 display the results of calibrating the camera with a tilt camera model that enforces equal ray angles in object and image space. In the middle part,  $s_x$  is fixed (enforcing square pixels), while  $s_y$  is optimized in the lower part. As predicted by Remark 13, if  $s_x$  is fixed, a large RMS error results if the lens is tilted in any direction. If  $s_x$  is optimized, the RMS error for the tilt around the vertical axis is much smaller (albeit not as small as for the proposed tilt camera model because the lens has distortions), while the RMS error of the diagonal tilt remains large. In both cases, a tilt angle  $\tau$  that deviates significantly from the nominal value is returned.

As a further example for SLR lenses, a Nikon AF-S Micro-Nikkor 105 mm f/2.8G was calibrated with a near focus setting, resulting in a pupil magnification factor of  $\approx 0.73$ . Furthermore, two standard machine vision lenses



**Fig. 18** The experimental setup used to obtain the results in Table 2. A Sigma 150 mm F2.8 macro lens and a Nikon D3 camera were attached to the Novoflex Balpro T/S bellows. The image also shows the calibration object

were tested. For a Cosmicar/Pentax C1614-M 16 mm lens, a pupil magnification factor of  $\approx 3$  was obtained, while for a Cosmicar B1214D-2 12.5 mm it was  $\approx 2.7$ . All of these SLR and machine vision lenses cannot be modeled correctly by camera models that force the ray angles in object and image space to be identical. A third machine vision lens, a Schneider Kreuznach BK-Makro CPN-S 50 mm lens, has a pupil magnification factor so close to 1 that it could be calibrated correctly with a traditional tilt camera model.

When a Nikon PC-E Nikkor 24 mm f/3.5D tilt/shift lens (the lens that is shown in the right image of Fig. 5) was mounted on a Nikon D3 camera and calibrated with the perspective tilt lens model, the image plane distance  $d$  consistently converged to values  $\geq 10^{60}$  m. Therefore, the lens was calibrated with the image-side telecentric tilt lens model. The tilt angle was adjusted to  $\approx 0^\circ$  and  $\approx 6^\circ$  using the tilt scale on the lens. The results are displayed in Table 3. From the top part of the table, we can see that the RMS errors increase somewhat if the lens is tilted. Nevertheless, the tilt angles are very consistent with the nominal tilt angles.

**Table 2** Calibration results with a Nikon D3 camera and a Sigma 150mm F2.8 macro lens attached to a Novoflex Balpro T/S bellows

	Proposed perspective tilt camera model		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 90^\circ$	$\tau \approx 6^\circ$ $\rho \approx 135^\circ$
RMS error (Pixel)	0.12552	0.14296	0.13522
$c$ (m)	0.429985	0.431056	0.429849
$K_1$ ( $m^{-2}$ )	-3.44217	-1.95776	-1.51355
$K_2$ ( $m^{-4}$ )	6224.19	1488.87	-706.09
$K_3$ ( $m^{-6}$ )	-4904890.6	1567851.4	3261860.5
$P_1$ ( $m^{-2}$ )	0.0067018	-0.0077062	-0.0021945
$P_2$ ( $m^{-2}$ )	-0.0008882	-0.0011449	-0.0040797
$d$ (m)	0.000765	0.129138	0.138867
$\tau$ ( $^\circ$ )	0.00351	5.73583	6.12792
$\rho$ ( $^\circ$ )	286.6215	87.0059	133.2228
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$c_x$ (Pixel)	2125.07	1554.51	1609.62
$c_y$ (Pixel)	1471.31	1479.47	968.19
	Camera model with equal ray angles, $s_x$ fixed		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 90^\circ$	$\tau \approx 6^\circ$ $\rho \approx 135^\circ$
RMS error (Pixel)	0.13177	0.66399	1.73470
$c$ (m)	0.431159	0.435400	0.422823
$K_1$ ( $m^{-2}$ )	-1.45076	1.20055	-5.65209
$K_2$ ( $m^{-4}$ )	-105.82	-1221.96	1041.01
$K_3$ ( $m^{-6}$ )	36541.1	159947.8	-91270.8
$P_1$ ( $m^{-2}$ )	-0.0206285	0.1662838	0.1190503
$P_2$ ( $m^{-2}$ )	-0.0048866	-0.0088579	0.1226469
$\tau$ ( $^\circ$ )	4.17771	17.74642	14.85163
$\rho$ ( $^\circ$ )	281.95401	88.68077	146.67340
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$c_x$ (Pixel)	4144.37	-2866.07	685.62
$c_y$ (Pixel)	1751.24	1427.25	-2005.61
	Camera model with equal ray angles, $s_x$ optimized		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 90^\circ$	$\tau \approx 6^\circ$ $\rho \approx 135^\circ$
RMS error (Pixel)	0.12649	0.20292	1.59371
$c$ (m)	0.430459	0.431696	0.425048
$K_1$ ( $m^{-2}$ )	-2.32917	-1.85163	-6.93979
$K_2$ ( $m^{-4}$ )	1017.87	934.69	3177.50
$K_3$ ( $m^{-6}$ )	-407686.6	1446855.5	-623427.6
$P_1$ ( $m^{-2}$ )	0.0055757	-0.0060708	0.1448057
$P_2$ ( $m^{-2}$ )	0.0229290	-0.0123213	0.0522046
$\tau$ ( $^\circ$ )	2.32281	18.67772	15.53254
$\rho$ ( $^\circ$ )	235.7823	89.30650	157.62229
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45957	8.86725	8.28404
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45000	8.45000	8.45000
$c_x$ (Pixel)	2179.88	1416.39	705.32
$c_y$ (Pixel)	-201.90	1544.09	-1411.09

**Table 3** Calibration results with a Nikon D3 camera and a Nikon PC-E Nikkor 24 mm f/3.5D tilt/shift lens

	Proposed image-side telecentric tilt camera model		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 270^\circ$	$\tau \approx 6^\circ$ $\rho \approx 330^\circ$
RMS error (Pixel)	0.10965	0.17466	0.17374
$c$ (m)	0.0275712	0.0275857	0.0275460
$K_1$ ( $m^{-2}$ )	85.04494	80.56032	82.35937
$K_2$ ( $m^{-4}$ )	5921.84	36041.03	21691.21
$K_3$ ( $m^{-6}$ )	4374052.8	-35282927.4	-12617450.7
$P_1$ ( $m^{-2}$ )	-0.0063425	0.0717010	0.0665242
$P_2$ ( $m^{-2}$ )	-0.0033489	-0.0059848	0.0317831
$\tau$ ( $^\circ$ )	0.76712	5.80991	5.86474
$\rho$ ( $^\circ$ )	29.1497	268.3920	299.2761
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$c_x$ (Pixel)	2139.80	1745.03	1808.66
$c_y$ (Pixel)	1406.30	1398.72	1198.77
	Camera model with equal ray angles, $s_x$ fixed		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 270^\circ$	$\tau \approx 6^\circ$ $\rho \approx 330^\circ$
RMS error (Pixel)	0.11391	0.43223	0.41050
$c$ (m)	0.0275301	0.0279162	0.0278926
$K_1$ ( $m^{-2}$ )	84.10718	93.29245	93.94938
$K_2$ ( $m^{-4}$ )	13265.11	-10120.40	-24810.57
$K_3$ ( $m^{-6}$ )	-12975174.6	-28823291.0	416994.8
$P_1$ ( $m^{-2}$ )	-0.0277049	0.4894341	0.4279702
$P_2$ ( $m^{-2}$ )	0.0451126	-0.0144357	0.2361730
$\tau$ ( $^\circ$ )	1.19386	9.37088	9.51047
$\rho$ ( $^\circ$ )	23.23408	268.82579	300.08426
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45	8.45	8.45
$c_x$ (Pixel)	2112.73	2293.13	2289.54
$c_y$ (Pixel)	1469.51	1387.42	1477.57
	Camera model with equal ray angles, $s_x$ optimized		
	$\tau \approx 0^\circ$	$\tau \approx 6^\circ$ $\rho \approx 270^\circ$	$\tau \approx 6^\circ$ $\rho \approx 330^\circ$
RMS error (Pixel)	0.11278	0.19790	0.40034
$c$ (m)	0.0275227	0.0276073	0.0281010
$K_1$ ( $m^{-2}$ )	84.15884	80.84468	96.92907
$K_2$ ( $m^{-4}$ )	12761.93	34023.77	-47844.93
$K_3$ ( $m^{-6}$ )	-12004974.3	-33947307.4	33119191.5
$P_1$ ( $m^{-2}$ )	-0.0331286	0.0565760	0.5362288
$P_2$ ( $m^{-2}$ )	0.0333186	-0.0357765	0.1826242
$\tau$ ( $^\circ$ )	1.01722	0.86247	11.55105
$\rho$ ( $^\circ$ )	35.24591	155.88407	288.36630
$s_x$ ( $\mu m$ Pixel $^{-1}$ )	8.45039	8.40596	8.47532
$s_y$ ( $\mu m$ Pixel $^{-1}$ )	8.45000	8.45000	8.45000
$c_x$ (Pixel)	2106.06	1723.56	2453.06
$c_y$ (Pixel)	1454.08	1354.72	1414.04



The middle and lower parts of Table 3 display the results of calibrating the lens with a camera model that enforces equal ray angles in object and image space. Again, as predicted by Remark 13, the RMS errors are large if the lens is tilted and  $s_x$  is fixed, while the error for the tilt around the vertical axis is smaller if  $s_x$  is optimized. As before, the tilt angles deviate significantly from their nominal values.

The fact that the Nikon 24mm tilt/shift lens is image-side telecentric piqued our interest and we calibrated further SLR tilt/shift lenses. It turned out that the Nikon PC-E Micro Nikkor 45 mm f/2.8D, the Nikon PC-E Micro Nikkor 85 mm f/2.8D, the Canon TS-E 17 mm f/4L, and the Canon TS-E 24 mm f/3.5L II lenses all are telecentric in image space.<sup>18</sup> Furthermore, a calibration of the standard machine vision lens Pentax C815B showed that this lens, too, is image-side telecentric. Again, none of these lenses can be calibrated correctly by a tilt camera model that forces rays to have the same angle in object and image space.

As a final example, Table 4 displays the results of calibrating two object-side telecentric lenses. The lenses were mounted onto the special camera housing described at the beginning of this section. Apart from facilitating tilts around the vertical axis, the camera housing also allows the position of the image plane to be adjusted, thereby facilitating a focusing of the lens to object planes with different distances.<sup>19</sup> The image plane was adjusted to a slightly larger distance than the nominal 17.526 mm of a standard C-mount lens to allow a large tilt of the lenses. Note that this increases the magnification of the lenses slightly. The camera uses an IDS uEye UI-1222LE sensor. The lenses were tilted around the vertical axis by the maximum angle the camera housing allowed ( $\approx 14^\circ$  for both experiments; because of different image plane positions, the angle differed between the experiments). The Vicotar T201/0.19 lens has a nominal magnification of 0.19. The nominal magnification of the V.S. Technologies L-VS-TC017 lens is 0.17. The calibrated tilt angles are close to the expected values. Furthermore, as expected, the calibrated magnifications are slightly larger than the nominal values. In this experiment, the division model was sufficient to model the lenses accurately.

## 12 Rectification of Stereo Images of a Perspective and a Telecentric Camera

With the results of the calibration approach discussed in Sects. 9 and 10, a stereo reconstruction can be performed.

<sup>18</sup> Unfortunately, we could not obtain the remaining SLR tilt/shift lenses, the Canon TS-E 45 mm f/2.8 and the Canon TS-E 90 mm f/2.8. However, it would not be far-fetched to conjecture that they also are telecentric in image space.

<sup>19</sup> Telecentric lenses usually have a fixed working distance.

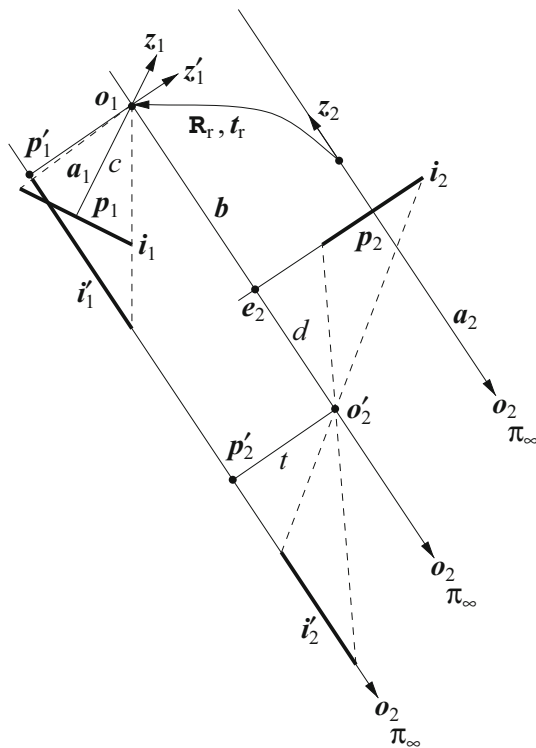
**Table 4** Calibration results with two object-side telecentric lenses: a Vicotar T201/0.19 and a V.S. Technologies L-VS-TC017 lens

	T201/0.19	L-VS-TC017
RMS error (Pixel)	0.03742	0.12973
$m$ (Scalar)	0.2157109	0.1977478
$\kappa$ ( $\text{m}^{-2}$ )	199.485	-2994.678
$d$ (m)	0.0432999	0.0298853
$\tau$ ( $^\circ$ )	15.11307	13.38796
$\rho$ ( $^\circ$ )	91.81762	92.27376
$s_x$ ( $\mu\text{m Pixel}^{-1}$ )	6.0	6.0
$s_y$ ( $\mu\text{m Pixel}^{-1}$ )	6.0	6.0
$c_x$ (Pixel)	135.79	200.73
$c_y$ (Pixel)	185.09	175.52

Stereo algorithms require that the images are rectified in such a manner that the epipolar lines in the rectified images are horizontal and at the same row coordinate. For pairs of perspective and pairs of telecentric cameras,<sup>20</sup> the geometry of the stereo rectification is well understood (Hartley and Zisserman 2003, Chapter 11.12). For a stereo image pair of a perspective and a telecentric camera, we could, in principle, use any of the homography-based algorithms that are capable of handling this configuration correctly, e.g., (Gluckman and Nayar 2001). However, these approaches are purely based on 2D projective transformations, which gives us no insight about the 3D geometry of the problem. We will show below that the rectification of a perspective and a telecentric stereo image pair will lead to a kind of object-side telecentric lens that we have not considered so far.

The stereo geometry of a perspective and a telecentric camera is shown in Fig. 19. Without loss of generality, we can assume that the perspective camera is the first camera and that it lies to the left of the second, telecentric, camera. In principle, the stereo geometry is analogous to that of two perspective cameras. The base  $\mathbf{b}$  connects the two projection centers  $\mathbf{o}_1$  and  $\mathbf{o}_2$ . The difference is that  $\mathbf{o}_2$  lies in the plane at infinity  $\pi_\infty$ . Therefore,  $\mathbf{b}$  is parallel to the optical axis  $\mathbf{a}_2$  of the telecentric camera. The two images must be projected onto a common image plane that is parallel to  $\mathbf{b}$ . Consequently, the rectified image plane, the base  $\mathbf{b}$ , and the optical axis  $\mathbf{a}_2$  all intersect in  $\mathbf{o}_2$ . The orientation of the rectified image plane has one degree of freedom: it can be rotated around  $\mathbf{b}$ . We select the normal of the rectified image plane to point from  $\mathbf{o}_1$  to the point on  $\mathbf{a}_2$  that has the shortest distance

<sup>20</sup> In this section, the terms perspective and telecentric refer to the object-side projection characteristics of the lenses. The image-side projection characteristics are immaterial for the discussion in this section. Everything we describe in this section works for lenses that are perspective or telecentric in image space, no matter whether they are tilted or not.



**Fig. 19** The geometry of rectifying a stereo image pair of a perspective and a telecentric camera. The perspective camera is described by its projection center  $o_1$ , its optical axis  $a_1$ , its image plane  $i_1$ , its principal point  $p_1$ , its principal distance  $c$ , and its camera coordinate system, of which only the  $z_1$  axis is shown. The telecentric camera is described by its projection center  $o_2$ , lying in the plane at infinity  $\pi_\infty$ , its optical axis  $a_2$ , its image plane  $i_2$ , its principal point  $p_2$ , and its camera coordinate system, of which only the  $z_2$  axis is shown. As described in Remark 4, the origin of the camera coordinate system lies at a finite location. The relative orientation of the telecentric camera with respect to the projective camera is given by  $R_r$  and  $t_r$ . The base of the stereo system is the line  $b$  that connects  $o_1$  and  $o_2$ . It intersects  $i_2$  in the epipole  $e_2$ . The base  $b$  is parallel to  $a_2$ . The rectifying image plane is parallel to  $b$  and to  $a_2$ . Its normal is given by  $z'_1$ . The rectifying image plane,  $a_2$ , and  $b$  all intersect in  $o_2$ . The rectified image of  $i_1$  is  $i'_1$ , that of  $i_2$  is  $i'_2$ . The rectification of the perspective camera image is performed by a standard rotation of the image plane around the projection center  $o_1$  and, possibly, a change of the principal distance. The new principal point is denoted by  $p'_1$ . To rectify  $i_2$ , a perspective projection of  $i_2$  onto  $i'_2$  must be performed. Its projection center  $o'_2$  must lie on  $b$ . The distance  $d$  of  $o'_2$  from  $e_2$  can be chosen arbitrarily. The principal point of  $i'_2$  is given by  $p'_2$ . The distance  $t$  of  $i'_1$  and  $i'_2$  from  $b$  can be chosen arbitrarily. Note that the rectified telecentric camera with principal point  $e_2$ , virtual image plane  $i_2$ , and image plane  $i'_2$  is an object-side telecentric camera with an image plane tilted by  $90^\circ$  and shifted by  $t$

from  $o_1$ . This results in smaller perspective distortions than other choices of the normal.

The rectification of the perspective camera is the usual rotation of the camera around  $o_1$  and, possibly, a change of the principal distance. The new viewing direction of the rectified camera is  $z'_1$ , which is also the normal of the rectified image plane. The axis  $x'_1$  of the rectified camera coordinate

system of the perspective camera (not shown in Fig. 19) must be parallel to  $b$ . The axis  $y'_1$  must be chosen as  $y'_1 = z'_1 \times x'_1$ .

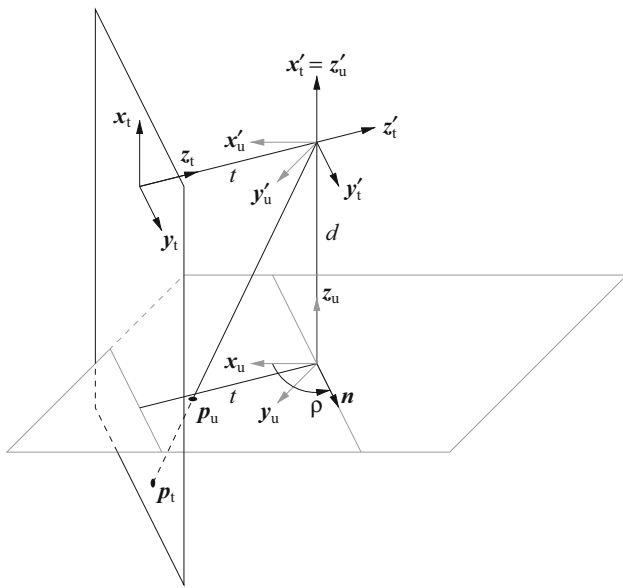
The rectification of the telecentric camera is more interesting. We must perspectively project the image  $i_2$  onto the rectified image plane, resulting in the image  $i'_2$ . Note that  $i'_2$  is orthogonal to  $i_2$ , i.e., is rotated by  $90^\circ$  with respect to  $i_2$ . Clearly, the perspective center  $o'_2$  of this projection must lie on the base  $b$  for the epipolar lines to become parallel and horizontal.<sup>21</sup> The intersection of  $b$  with  $i_2$  is the epipole  $e_2$ . The distance  $d$  of  $o'_2$  from  $e_2$  is a parameter that can be chosen freely. Small values of  $d$  lead to narrow rectified images, large values of  $d$  to wide images. We choose  $d$  such that  $i'_2$  has the same width as  $i'_1$  to cause the perspective distortions in both rectified images to be approximately equal. The distance  $t$  of the rectified images planes to the base is another free parameter. We simply set it to  $c$ , i.e., to the principal distance of the perspective camera. Other choices of  $t > 0$  are possible, but would simply scale the rectified images. Note that the rectified telecentric camera with principal point  $e_2$ , virtual image plane  $i_2$ , and image plane  $i'_2$  is an object-side telecentric camera with an image plane tilted by  $90^\circ$  and shifted by  $t$ . Its optical axis  $b$  is shifted from the optical axis of the original telecentric camera by the vector  $e_2 - p_2$ , where  $p_2$  is the principal point of the original camera.

The camera model of the rectified telecentric camera cannot be represented by the model for object-side telecentric tilt cameras described in Sect. 7. The model of Sect. 7 describes the tilted image plane by three parameters in an affine parameterization. This is analogous to describing the tilted image plane by  $z = ax + by + d$ . Obviously, this parameterization can represent all planes that are not perpendicular to the plane  $z = 0$ , while it cannot represent planes that are perpendicular to  $z = 0$ . Therefore, we must extend the model for tilt homographies to handle image planes that are tilted by  $90^\circ$  and shifted by  $t$  with respect to the original principal point.<sup>22</sup>

The geometry of projecting a point from an untilted image plane to an image plane that is tilted by  $90^\circ$  is shown in Fig. 20. Unlike for tilts smaller than  $90^\circ$ , the origin of the tilted coordinate system  $(x_t, y_t, z_t)$  does not coincide with that of the untilted coordinate system  $(x_u, y_u, z_u)$  because the optical axis intersects the tilted image plane at infinity. Therefore, we use the traditional principal point as the reference point of the tilted image plane. The geometry is equivalent to a camera that rotates around the projection center and changes its principal distance from  $d$  to  $t$ . Thus, the projection can be modeled by the homography

<sup>21</sup> The projection center must lie in the plane spanned by  $b$  and  $y'_1$  for the epipolar lines to become parallel. Furthermore, it must lie in the plane spanned by  $b$  and  $z'_1$  for the epipolar lines to become horizontal. Therefore, it must lie on  $b$ .

<sup>22</sup> As noted above, the translation by  $t$  can be regarded as a rotation around  $o_2$  in the plane at infinity  $\pi_\infty$ .



**Fig. 20** The projection of a point  $p_u$  from the untitled image plane to a point  $p_t$  in an image plane that is tilted by  $90^\circ$ . The camera is perspective in image space. The coordinate system of the untitled image plane is given by  $(x_u, y_u, z_u)$ , that of the tilted image plane by  $(x_t, y_t, z_t)$ . The untitled camera’s viewing direction is along the  $z_u$  axis, which points towards the scene. The vector  $\mathbf{n}$ , which forms an angle  $\rho$  with  $\mathbf{x}_u$ , denotes the direction along which the two image planes intersect. The tilted image plane is shifted by the distance  $t$  in the direction perpendicular to  $\mathbf{n}$ . These two coordinate systems can also be attached to the image-side projection center as  $(x'_u, y'_u, z'_u)$  and  $(x'_t, y'_t, z'_t)$ . The distance from the projection center to the intersection of the optical axis with the untitled image plane is  $d$

$$\mathbf{H}_{90^\circ} = \mathbf{K}_t \mathbf{R}^\top \mathbf{K}_u^{-1}, \tag{59}$$

where  $\mathbf{K}_u$  is given by (26) and  $\mathbf{K}_t$  is given by

$$\mathbf{K}_t = \begin{pmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{60}$$

The rotation matrix  $\mathbf{R}$  is given by expressing  $(x_t, y_t, z_t)$  in terms of  $(x_u, y_u, z_u)$ . Since  $z_u$  is the base of the stereo system, we obviously must choose  $x_t$  parallel to  $z_u$ . One choice is to set  $x_t = z_u = (0, 0, 1)^\top$ . As shown in Fig. 20, we can set  $y_t = \mathbf{n}$ , where  $\mathbf{n}$  is given by (21). Finally, we can set  $z_t = \mathbf{n}^\perp = (-\sin \rho, \cos \rho, 0)^\top$ . A second choice is given by  $x_t = -z_u, y_t = -\mathbf{n}$ , and  $z_t = \mathbf{n}^\perp$ . Substituting the two choices into (59), we obtain two possible homographies:

$$\mathbf{H}_{90^\circ} = \begin{pmatrix} 0 & 0 & t \\ \frac{t}{d} \cos \rho & \frac{t}{d} \sin \rho & 0 \\ -\frac{1}{d} \sin \rho & \frac{1}{d} \cos \rho & 0 \end{pmatrix} \tag{61}$$

and

$$\mathbf{H}_{90^\circ} = \begin{pmatrix} 0 & 0 & -t \\ -\frac{t}{d} \cos \rho & -\frac{t}{d} \sin \rho & 0 \\ -\frac{1}{d} \sin \rho & \frac{1}{d} \cos \rho & 0 \end{pmatrix} \tag{62}$$

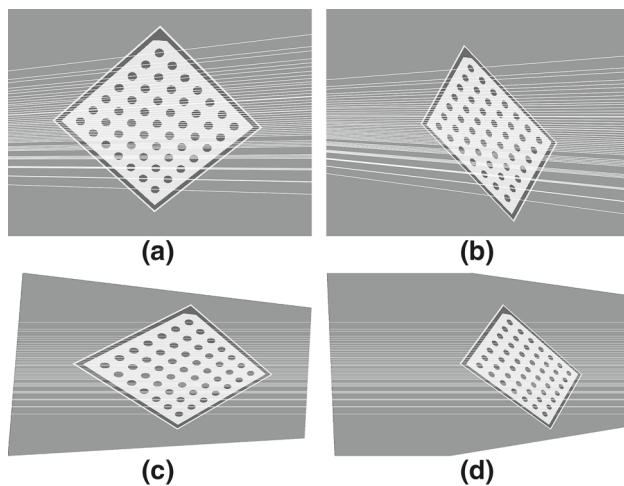
The appropriate choice is (61) if the telecentric camera is to the left of the perspective camera and (62) otherwise. We also note that (62), for  $t < 0$ , is identical to (61), for  $t > 0$ , and vice versa. Therefore, we only need either (61) or (62) and can select the appropriate solution via the sign of  $t$ . The vector  $\mathbf{n}$  is given by the projection of  $z'_1$  into the undistorted image plane of the telecentric camera, rotated by  $-\pi/2$ .

With this, a point in the image plane  $i_2$  is projected into the rectified image plane  $i'_2$  by the following transformations. First, the point is transformed into the undistorted image plane of the telecentric camera by the inverse of (13), optionally the inverse of (24) or (30), depending on the camera model, and (10) or (12), depending on the distortion model. Next, the point is translated by the offset of the epipole  $e_2$ . Let the coordinates of the epipole in the undistorted image plane be denoted by  $(e_x, e_y)^\top$ . Since the coordinates of the principal point  $p_2$  in the undistorted image plane are  $(0, 0)^\top$ , the translation vector is  $(-e_x, -e_y)^\top$ . The translated point is then projected into the rectified image plane by (61) or (62). Finally, the projected point is transformed into the image coordinate system of the rectified image by (13). To transform points from  $i'_2$  into  $i_2$ , which is required for image rectification, the inverses of the above transformations must be applied in the reverse order.

There are some free parameters in the above transformations that still must be determined. First, to prevent the rectified images from becoming too wide, we select  $s_x$  and  $s_y$  such that the pixels in the rectified images are approximately the same size as the pixels in the original images. This is done by projecting a square of size 1 in the center of both rectified images back to the original images and using the inverse of the average of the resulting side lengths of the transformed squares as the scaling factors for  $s_x$  and  $s_y$ , respectively. Next, we select  $c_y$  of the two cameras equal based on the intersection of the bounding boxes of the two rectified images to ensure that the epipolar lines have the same row coordinate. Finally, the values of  $c_x$  are determined based on the individual bounding boxes of the two rectified images.

A synthetic example for the rectification of a stereo image pair of a perspective and a telecentric camera is shown in Fig. 21. The cameras are tilted by an angle of  $45^\circ$  with respect to each other. To show the epipolar lines for selected points before and after rectification, we use simulated images of a calibration object that we used previously (Steger et al. 2008, Chapter 3.9). This calibration object has fewer control points than the calibration object shown in Fig. 17, which allows us to display the epipolar lines corresponding to the centers of the control points. The images were created with synthetic camera parameters, i.e., the images in Fig. 21 are not used for calibration. They merely serve to have clearly identifiable points (the centers of the control points) for which the epipolar lines can be displayed. Figure 21 shows that the epipolar

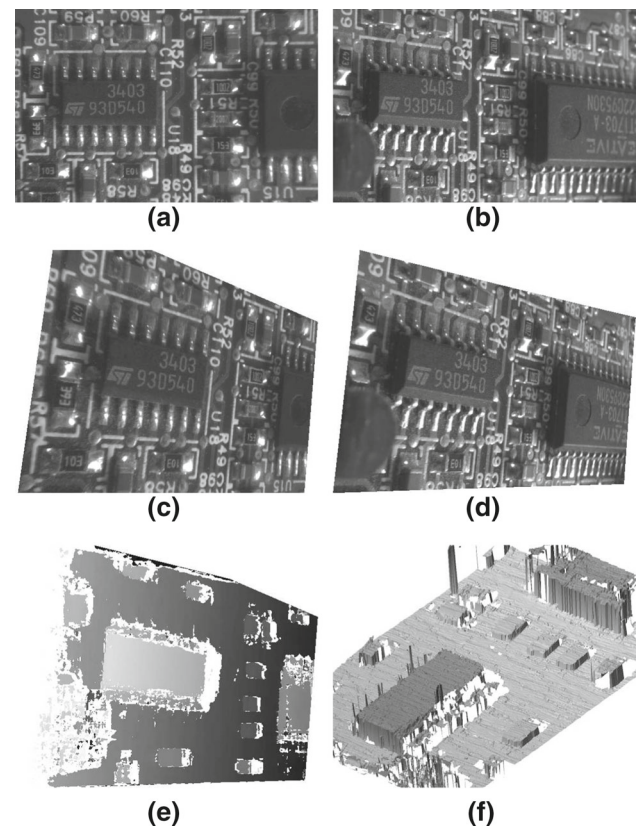




**Fig. 21** A synthetic example for the rectification of a stereo image pair of a perspective and a telecentric camera. The image of the perspective camera, along with the epipolar lines through the centers of the control points, is shown in **a**, the image of the telecentric camera is shown in **b**. The corresponding rectified images and epipolar lines are shown in **c**, **d**

lines of the original images are slanted and those of the rectified images are horizontal and at the same row coordinates, illustrating that the rectification algorithm described above works correctly.

A real example for the rectification of a stereo image pair of a telecentric and a perspective camera and for the corresponding 3D reconstruction is shown in Fig. 22. Both cameras were equipped with an IDS uEye UI-1222LE sensor. The telecentric camera was not tilted and used a V.S. Technologies L-VS-TC017 lens. The perspective camera used the special housing, described in Sect. 11, that allows the lens to be tilted around the vertical axis. The camera was equipped with a Cosmicar B1214D-2 12.5 mm lens with a 2 mm extension tube, tilted by  $\approx 3.5^\circ$ . The stereo setup was calibrated with the approach described in Sect. 9. The calibrated relative pose of the perspective camera with respect to the telecentric camera was computed as  $t_x = -46.00$  mm,  $t_y = 6.41$  mm,  $t_z = 19.23$  mm,  $\alpha = 3.38^\circ$ ,  $\beta = 37.39^\circ$ , and  $\gamma = -2.66^\circ$ . Figure 22a, b shows two images of a printed circuit board (PCB) acquired with this setup. The telecentric camera looks perpendicularly onto the PCB, the perspective camera at an angle of  $\approx 37^\circ$ . Figure 22c, d displays the rectified images. To show that the calibration and rectification worked correctly, the stereo disparities computed from the two rectified images are shown in Fig. 22e. A perspective view of the metric 3D reconstruction is displayed in Fig. 22f. Note that all components on the PCB (ICs, resistors, and capacitors) have been reconstructed correctly. The surfaces of the components and the PCB itself are perfectly flat. There are a few mismatches that are caused by specular reflections, mainly on the leads of the IC, and by the large occlusion in the



**Fig. 22** A real example for the rectification of a stereo image pair of a telecentric and a perspective camera and for the corresponding 3D reconstruction. **a** Image of the telecentric camera, which looks perpendicularly at a PCB. **b** Image of the perspective tilt camera, which looks at the PCB at an angle of  $\approx 37^\circ$ . **c** Rectified telecentric image. **d** Rectified perspective image. **e** Computed stereo disparities. **f** Visualization of the computed metric 3D reconstruction. For better visibility, the reconstructed surface has been colored according to the  $z$  value of the reconstruction in the camera coordinate system of the telecentric camera. Darker surface colors correspond to points that are closer to the telecentric camera

lower left part of the perspective image. Overall, the reconstruction is of very high quality, showing that the calibration and rectification work correctly.

*Remark 21* It might appear that (61) and (62) lead to a completely new camera model for object-side telecentric tilt cameras. However, this is not the case. If (61) or (62) are inserted into (54), a camera matrix  $\mathbf{M}$  that lies in the set  $D$  of Lemma 2 in Appendix A.3 is obtained. Computing the DIAC of such a camera matrix  $\mathbf{M}$  results in:

$$\omega_{\mathbf{H}_{90^\circ}}^* = \begin{pmatrix} \frac{c_x^2}{d^2} & \frac{c_x c_y}{d^2} & \frac{c_x}{d^2} \\ \frac{c_x c_y}{d^2} & \frac{c_y^2 + a_y^2 t^2}{d^2} & \frac{c_y}{d^2} \\ \frac{c_x}{d^2} & \frac{c_y}{d^2} & \frac{1}{d^2} \end{pmatrix}. \quad (63)$$

The DIAC is independent of  $\rho$  and tautologically fulfills equations (105) and (106). Equation (107), however, evaluates to  $a_y^2 t^2 / d^4$  and, therefore, is only fulfilled if  $a_y =$



$0, t = 0$ , or  $d = \infty$ . These cases, however, cannot occur since we must require  $t \neq 0$  and  $d < \infty$  for  $\mathbf{H}_{90^\circ}$  and  $a_y \neq 0$  for  $\mathbf{K}$ , and therefore  $\mathbf{M}$ , to be regular. Thus, the decomposition result from Lemma 2 can be applied. In particular, the matrix  $\mathbf{M}$  that arises from the rectification can be interpreted as an object-side telecentric tilt camera with  $\cos \rho = 0$ .

As an example, consider an object-side telecentric tilt camera for rectification, using (62) as the tilt homography, with interior orientation parameters  $m = 1, d = 0.02, t = 0.015, \rho = 300^\circ, s_x = s_y = 2 \times 10^{-5}, (c_x, c_y)^\top = (1500, 1000)^\top$ , and a pose given by  $\alpha = 30^\circ, \beta = -20^\circ, \gamma = 10^\circ$ , and  $(t_x, t_y, t_z)^\top = (0.158, 0.078, 0)^\top$ . By the decomposition in the proof of Lemma 2, this camera can be represented by a regular object-side telecentric tilt camera with parameters  $m = 1, d = 0.01118034, \tau = 48.189685^\circ, \rho = 90^\circ, s_x = 1 \times 10^{-5}, s_y = 1.333333 \times 10^{-5}, (c_x, c_y)^\top = (0, 1000)^\top$ , and a pose given by  $\alpha = 16.164849^\circ, \beta = -32.081247^\circ, \gamma = -20.733967^\circ$ , and  $(t_x, t_y, t_z)^\top = (0.165832, -0.011450, 0)^\top$ . Needless to say, the parameters of the regular object-side telecentric tilt camera are extremely unintuitive. Therefore, while Theorem 3 and, in particular, Lemma 2 show that no special camera model for object-side telecentric tilt cameras for rectification would be required, it is preferable to have an explicit model for rectification because this facilitates an intuitive interpretation of the geometric parameters.

*Remark 22* The above discussion shows that the rectified telecentric camera can be regarded as an object-side telecentric tilt camera. Obviously, the rectified perspective camera can also be regarded as a perspective tilt camera. This also applies in the case of two perspective cameras. For a stereo system with two telecentric cameras, the images also must be projected onto a common image plane. This projection can be represented by rotations around the optical axis of the two cameras since the only property that must be fulfilled is that the lines of the images are parallel to the direction that connects the two projection centers in the plane at infinity. The fact that, seemingly, there is no tilt involved is a consequence of Proposition 6. The tilt that would occur in the projection onto the common image plane can be compensated by changing  $s_x$  appropriately. Since  $s_x$  can be chosen freely anyway, the rectification by a rotation around the optical axis effectively is a tilt that is being undone by virtually selecting a smaller  $s_x$ . This shows that, for all combinations of camera types, stereo rectification involves cameras with tilt lenses in a very natural manner.

## 13 Conclusions

We have proposed models for cameras with tilt lenses that correctly model the imaging geometry of lenses for which

the ray angles in object and image space differ. The models cover all lens types that are in common use for machine vision and consumer cameras: entocentric, image-side telecentric, object-side telecentric, and bilateral telecentric lenses. We have shown that the tilt can be modeled by orthographic or perspective homographies that project points from an untilted virtual image plane to the tilted image plane. An important aspect of the parameterization of the tilt homographies is that their parameters are easy to understand for the user. Furthermore, we have analyzed the degeneracies of the proposed camera models and have described how they can be handled automatically or manually by the user. The analysis of the degeneracies also has led to theorems that show that all finite projective cameras, affine cameras, and general cameras at infinity, i.e., all projective cameras, can be regarded as cameras with tilt lenses. Each theorem provides a minimal parameterization of the respective class of camera matrices. In addition, we have proposed two efficient and accurate algorithms to remove the perspective and distortion bias from the positions of circular control points, and thus to increase the accuracy of the calibration results. The proposed algorithms are at least an order of magnitude faster than the existing approaches that are capable of compensating perspective and distortion bias. Experiments have verified that the proposed camera models achieve accurate calibration results on numerous configurations of cameras and tilt lenses. The experiments also have shown that the camera models that were proposed up to now are unable to model many existing lenses correctly. Finally, we have described a geometric algorithm to rectify a stereo image pair of a perspective and a telecentric camera. This has led us to an interesting subclass of the tilt camera models: an object-side telecentric camera with an image plane that is tilted by  $90^\circ$ .

**Acknowledgements** I am grateful to Rebecca König for working out the proof of Theorem 3.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix

### A.1 Proof of Theorem 1

An affine camera is represented by a camera matrix of the form

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (64)$$

for which the top left  $2 \times 3$  submatrix has rank 2 (Hartley and Zisserman 2003, Chapter 6.3.4).

Since we assume square pixels and the principal point at  $(0, 0)^T$ , the calibration matrix in (55) is given by

$$\mathbf{K} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{65}$$

If we denote the elements of the tilt homography  $\mathbf{H}_o$  in (31) by  $h_{ij}$ , the elements of the rotation matrix in  $\mathbf{H}_w$  by  $r_{ij}$ , and the translation vector in  $\mathbf{H}_w$  by  $(t_x, t_y, t_z)^T$ , (55) results in the following camera matrix

$$\mathbf{P}_{oo} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{66}$$

where

$$p_{ij} = a(h_{i1}r_{1j} + h_{i2}r_{2j}) \tag{67}$$

$$p_{i4} = a(h_{i1}t_x + h_{i2}t_y) \tag{68}$$

for  $i \in \{1, 2\}$  and  $j \in \{1, 2, 3\}$ .

To prove the theorem, we will show  $\mathbf{P}_{oo} = \mathbf{M}$  for an arbitrary affine camera matrix  $\mathbf{M}$ , i.e., we will show that every  $\mathbf{M}$  can be uniquely decomposed into the parameters of  $\mathbf{P}_{oo}$ . To do so, we first remove the pose parameters from  $\mathbf{P}_{oo}$  and  $\mathbf{M}$  by computing their DIACs. Let us denote the DIAC of  $\mathbf{M}$  by  $\omega_M^*$  and the DIAC of the camera matrix (66) by  $\omega_P^*$ . We must show that

$$\omega_P^* = \omega_M^*. \tag{69}$$

If we denote the elements of  $\omega_M^*$  by  $\omega_{ij}$ , (69) results in the following three equations, as a result of applying (57) to (66):

$$\omega_{11} = -\frac{a^2 \cos^2 \rho \sin^2 \tau - a^2}{\cos^2 \tau} \tag{70}$$

$$\omega_{22} = -\frac{a^2 \sin^2 \rho \sin^2 \tau - a^2}{\cos^2 \tau} \tag{71}$$

$$\omega_{12} = -\frac{a^2 \cos \rho \sin \rho \sin^2 \tau}{\cos^2 \tau}, \tag{72}$$

where  $\cos^2 \theta$  and  $\sin^2 \theta$  are abbreviations for  $(\cos \theta)^2$  and  $(\sin \theta)^2$ , respectively.

We can solve (72) for  $a^2$ :

$$a^2 = -\frac{\omega_{12} \cos^2 \tau}{\cos \rho \sin \rho \sin^2 \tau}. \tag{73}$$

Substituting  $a^2$  into (70) and (71) and simplifying yields

$$\omega_{12}(\cos^2 \rho \sin^2 \tau - 1) - \omega_{11} \cos \rho \sin \rho \sin^2 \tau = 0 \tag{74}$$

$$\omega_{12}(\sin^2 \rho \sin^2 \tau - 1) - \omega_{22} \cos \rho \sin \rho \sin^2 \tau = 0. \tag{75}$$

By subtracting (75) from (74), we obtain

$$\omega_{12}(1 - 2 \sin^2 \rho) \sin^2 \tau + (\omega_{22} - \omega_{11}) \cos \rho \sin \rho \sin^2 \tau = 0. \tag{76}$$

Let us assume for the moment that  $\sin \tau \neq 0$ . Then,

$$\omega_{12}(1 - 2 \sin^2 \rho) + (\omega_{22} - \omega_{11}) \cos \rho \sin \rho = 0. \tag{77}$$

This can be transformed to

$$2\omega_{12} \cos(2\rho) - (\omega_{11} - \omega_{22}) \sin(2\rho) = 0, \tag{78}$$

i.e.,

$$\tan(2\rho) = \frac{2\omega_{12}}{\omega_{11} - \omega_{22}}. \tag{79}$$

To handle the case  $\sin \tau = 0$ , we note that if  $\sin \tau = 0$ ,  $\omega_{12} = 0$  and  $\omega_{11} = \omega_{22} = a^2$ . Conversely, if  $\omega_{12} = 0$  and  $\omega_{11} = \omega_{22}$ , (72) can be transformed to  $a^2 \sin(2\rho) \sin^2 \tau = 0$  and the result of subtracting (71) from (70) can be transformed to  $a^2 \cos(2\rho) \sin^2 \tau = 0$ . Since  $\cos(2\rho)$  and  $\sin(2\rho)$  are never simultaneously 0, we must either have  $a = 0$  or  $\sin \tau = 0$ . Since we assume that  $\mathbf{M}$  is regular,  $a = 0$  cannot occur. Hence,  $\sin \tau$  must be 0. Consequently, if  $\omega_{12} = 0$  and  $\omega_{11} = \omega_{22}$ , we cannot determine  $\rho$  uniquely and can set it to 0. Otherwise, we can use (79) to find a solution for  $\rho$ . The special case can even be subsumed into the general case if we solve (79) for  $\rho$  by the two-argument arctangent function

$$\rho = \frac{1}{2} \operatorname{atan2}(2\omega_{12}, \omega_{11} - \omega_{22}) \tag{80}$$

since  $\operatorname{atan2}(0, 0) = 0$  by definition. A second possible solution is given by

$$\rho = \frac{1}{2} \operatorname{atan2}(-2\omega_{12}, \omega_{22} - \omega_{11}). \tag{81}$$

We will show below that (81) is always the correct solution.

Next, we add (74) and (75) and obtain

$$\left( \omega_{12} - \frac{(\omega_{11} + \omega_{22})}{2} \sin(2\rho) \right) \sin^2 \tau - 2\omega_{12} = 0. \tag{82}$$

Substituting  $\rho$  from (81) into (82) and using  $\sin(\operatorname{atan2}(\pm y, \pm x)) = \pm y/\sqrt{x^2 + y^2}$ , we obtain

$$\left( \omega_{12} + \frac{(\omega_{11} + \omega_{22})\omega_{12}}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}} \right) \sin^2 \tau - 2\omega_{12} = 0. \tag{83}$$

Solving for  $\sin^2 \tau$  results in

$$\sin^2 \tau = \frac{2\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2} + \omega_{11} + \omega_{22}}. \tag{84}$$

The right hand side of (84) is obviously  $\geq 0$ . Furthermore, we can tentatively assume

$$\frac{2\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2} + \omega_{11} + \omega_{22}} \leq 1. \tag{85}$$

Simplifying (85) algebraically shows that this is equivalent to  $\omega_{12}^2 \leq \omega_{11}\omega_{22}$ . This is nothing more than the Cauchy–Schwarz inequality. Therefore, the right hand side of (84) is  $\leq 1$  and (84) can always be solved for  $\tau$ :

$$\tau = \arcsin \left( \frac{2\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2} + \omega_{11} + \omega_{22}} \right). \tag{86}$$

To prove that (80) is always the incorrect solution for  $\rho$ , we substitute it into (82) and proceed as before to obtain

$$\sin^2 \tau = \frac{2\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2} - \omega_{11} - \omega_{22}}. \tag{87}$$

The right hand sides of (84) and (87) have different signs if their product is negative. Computing this product results in

$$\frac{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2}{\omega_{12}^2 - \omega_{11}\omega_{22}}. \tag{88}$$

The nominator is obviously  $\geq 0$ . If it is 0, we have  $\sin \tau = 0$  (see the discussion above), i.e., (84) and (87) return the same solution. Therefore, we can assume that the nominator is  $> 0$ . Because of the Cauchy–Schwarz inequality, the denominator is  $\leq 0$ . If the denominator were 0,  $\mathbf{M}$  would be singular since the first row of the left  $3 \times 3$  submatrix of  $\mathbf{M}$  would depend linearly on the second row. Consequently, the denominator is always  $< 0$  and therefore (87) has no solution except when  $\sin \tau = 0$ . This shows that (86) is the only correct solution for  $\tau$ .

Finally, we can substitute (81) and (86) into (73) and simplify to obtain

$$a = \sqrt{\frac{2(\omega_{11}\omega_{22} - \omega_{12}^2)}{\sqrt{(\omega_{22} - \omega_{11})^2 + 4\omega_{12}^2} + \omega_{11} + \omega_{22}}}. \tag{89}$$

Equations (81), (86), and (89) show that the interior orientation of the camera, i.e., the matrices  $\mathbf{K}$  (see (65)) and  $\mathbf{H}_o$  (see (31)), can be determined uniquely.

To determine the pose parameters, we can multiply  $\mathbf{M}$  from the left by  $\mathbf{H}_o^{-1}\mathbf{K}^{-1}$ . This results in

$$\mathbf{H}_o^{-1}\mathbf{K}^{-1}\mathbf{M} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{90}$$

The missing third row of the rotation matrix can be reconstructed by the vector product of the first two rows. This matrix can then be converted into any convenient rotation parameterization, e.g., Euler angles. Since the orthographic projection in object space (8), and therefore also (68), does not depend on  $t_z$ , we can set  $t_z$  to any value, e.g., to 0.  $\square$

### A.2 Proof of Theorem 2

A finite projective camera is represented by an arbitrary homogeneous  $3 \times 4$  matrix

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \tag{91}$$

for which the left hand  $3 \times 3$  submatrix  $\mathbf{M}_3$  is non-singular (Hartley and Zisserman 2003, Chapter 6.1). We assume  $\det(\mathbf{M}_3) > 0$ . If not, we multiply  $\mathbf{M}$  by  $-1$ .

Since we assume square pixels, the calibration matrix in (53) is given by

$$\mathbf{K} = \begin{pmatrix} a & 0 & c_x \\ 0 & a & c_y \\ 0 & 0 & 1 \end{pmatrix}. \tag{92}$$

If we denote the elements of the tilt homography  $\mathbf{H}_o$  in (31) by  $h_{ij}$ , the elements of the rotation matrix in  $\mathbf{H}_w$  by  $r_{ij}$ , and the translation vector in  $\mathbf{H}_w$  by  $(t_x, t_y, t_z)^T$ , (53) results in the following camera matrix

$$\mathbf{P}_{po} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}, \tag{93}$$

where

$$p_{1j} = a(h_{11}r_{1j} + h_{12}r_{2j}) + c_x r_{3j} \tag{94}$$

$$p_{2j} = a(h_{21}r_{1j} + h_{22}r_{2j}) + c_y r_{3j} \tag{95}$$

$$p_{3j} = r_{3j} \tag{96}$$

$$p_{14} = a(h_{11}t_x + h_{12}t_y) + c_x t_z \tag{97}$$

$$p_{24} = a(h_{21}t_x + h_{22}t_y) + c_y t_z \tag{98}$$

$$p_{34} = t_z \tag{99}$$

for  $j \in \{1, 2, 3\}$ .

To prove the theorem, we will show  $s\mathbf{P}_{po} = \mathbf{M}$  for an arbitrary finite projective camera matrix  $\mathbf{M}$ , i.e., we will show that every  $\mathbf{M}$  can be uniquely decomposed into the parameters of  $\mathbf{P}_{po}$ . Here,  $s \neq 0$  is a scaling factor that accounts for the fact that  $\mathbf{M}$  is homogeneous.

Since  $(r_{31}, r_{32}, r_{33})^\top$  has length 1,  $s$  can be computed as

$$s = \sqrt{m_{31}^2 + m_{32}^2 + m_{33}^2}. \tag{100}$$

In the following, we assume that  $\mathbf{M}$  has been normalized by dividing it by  $s$ . We then have

$$t_z = m_{34}. \tag{101}$$

Next, we compute the DIACs of  $\mathbf{P}_{po}$  and the normalized  $\mathbf{M}$  and require  $\omega_p^* = \omega_M^*$ . If we denote the elements of  $\omega_M^*$  by  $\omega_{ij}$ , we obtain

$$(c_x, c_y)^\top = (\omega_{13}, \omega_{23})^\top. \tag{102}$$

Hence, we have determined three of the eleven camera parameters:  $c_x$ ,  $c_y$ , and  $t_z$ .

We can now normalize  $\mathbf{P}_{po}$  and  $\mathbf{M}$  to have a principal point of  $(0, 0)^\top$  by multiplying them on the left by

$$\mathbf{K}_c = \begin{pmatrix} 1 & 0 & -c_x \\ 0 & 1 & -c_y \\ 0 & 0 & 1 \end{pmatrix}. \tag{103}$$

The matrix  $\mathbf{K}_c\mathbf{P}_{po}$  has the form

$$\mathbf{N}_{po} = \mathbf{K}_c\mathbf{P}_{po} = \begin{pmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \end{pmatrix}, \tag{104}$$

where  $n_{ij} = a(h_{i1}r_{1j} + h_{i2}r_{2j})$ ,  $j \in \{1, 2, 3\}$ , and  $n_{i4} = a(h_{i1}t_x + h_{i2}t_y)$  for  $i \in \{1, 2\}$ . Therefore, the first two rows of  $\mathbf{N}_{po}$  have the same structure as (66). Consequently, we can use the same approach as in Appendix A.1 to compute the remaining eight camera parameters.  $\square$

### A.3 Proof of Theorem 3

We will split the proof of Theorem 3 into two parts. As already shown in Proposition 5, the perspective tilt homographies that are used in the model for general cameras at infinity exhibit certain degeneracies that must be taken into account. Therefore, we will provide separate lemmas and proofs for the non-degenerate and degenerate cases. In the following, we denote the space of general cameras at infinity by  $S_M$ .

**Lemma 1** Every  $\mathbf{M} \in S_M^* = S_M \setminus D$ , where  $D$  is a set of measure 0, can be regarded as an object-side telecentric tilt camera. Here, a matrix  $\mathbf{M} \in S_M$  lies in  $D$  if the DIAC  $\omega_M^*$  of  $\mathbf{M}$  with elements  $\omega_{ij}$  satisfies at least one of the following equations

$$\omega_{12}\omega_{33} - \omega_{13}\omega_{23} = 0, \tag{105}$$

$$\omega_{11}\omega_{33} - \omega_{13}^2 = 0, \tag{106}$$

$$\omega_{22}\omega_{33} - \omega_{23}^2 = 0. \tag{107}$$

*Proof* We will prove this lemma by giving a set of maps  $\mathcal{A} = \{\phi_i\}_{i \in I}$ , where  $I$  is a finite set of indices, that maps every matrix  $\mathbf{M} \in S_M^*$  uniquely to the parameters of an object-side telecentric tilt camera. Here, the maps  $\phi_i$  are homeomorphisms defined on  $U_i \subset S_M^*$ , with  $\bigcup_{i \in I} U_i = S_M^*$ .

First, note that  $D$  is indeed a set of measure 0. This follows from the implicit function theorem applied to the function defined by the left hand side of each of the Eqs. (105)–(107). It states that one of the entries  $m_{ij}$  of  $\mathbf{M}$  can be expressed by the other entries  $m_{ij}$  involved in the equation to obtain the matrices that fulfill that equation. Now, this set of matrices is exactly the set we denoted by  $D$  and it has one dimension less than  $S_M$ . Thus,  $\ell(D) = 0$ , where  $\ell(\cdot)$  denotes the Lebesgue measure on  $\mathbb{R}^n$  with  $n = \dim(S_M)$ .

A general camera at infinity is represented by an arbitrary homogeneous  $3 \times 4$  matrix

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \tag{108}$$

for which the left hand  $3 \times 3$  submatrix  $\mathbf{M}_3$  is singular, i.e., for which  $\det(\mathbf{M}_3) = 0$  (Hartley and Zisserman 2003, Chapter 6.3.6).

If we denote the elements of the tilt homography  $\mathbf{H}_p$  in (32) by  $h_{ij}$ , the elements of the rotation matrix in  $\mathbf{H}_w$  by  $r_{ij}$ , and the translation vector in  $\mathbf{H}_w$  by  $(t_x, t_y, t_z)^\top$ , (54) results in the following camera matrix

$$\mathbf{P}_{op} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}, \tag{109}$$

where

$$p_{1j} = (h_{11}r_{1j} + h_{12}r_{2j})a_x + (h_{31}r_{1j} + h_{32}r_{2j})c_x \tag{110}$$

$$p_{2j} = (h_{21}r_{1j} + h_{22}r_{2j})a_y + (h_{31}r_{1j} + h_{32}r_{2j})c_y \tag{111}$$

$$p_{3j} = h_{31}r_{1j} + h_{32}r_{2j} \tag{112}$$

$$p_{14} = (h_{11}t_x + h_{12}t_y)a_x + (h_{31}t_x + h_{32}t_y + h_{33})c_x \tag{113}$$

$$p_{24} = (h_{21}t_x + h_{22}t_y)a_y + (h_{31}t_x + h_{32}t_y + h_{33})c_y \tag{114}$$

$$p_{34} = h_{31}t_x + h_{32}t_y + h_{33} \tag{115}$$

for  $j \in \{1, 2, 3\}$ .



Because the matrix  $\mathbf{M}$  that represents the general camera at infinity is only defined up to scale, we can assume that  $|m_{ij}| \leq 1, \forall i, j$ . If not, we scale  $\mathbf{M}$  by a suitable factor. For the elements  $\omega_{ij}$  of the DIAC  $\omega_M^*$  of  $\mathbf{M}$ , it follows that  $|\omega_{ij}| \leq 3, \forall i, j$ .

We will now show that  $s\mathbf{P}_{op} = \mathbf{M}$  ( $s > 0$ ). To do so, we first remove the pose parameters from  $\mathbf{P}_{op}$  and  $\mathbf{M}$  by computing their DIACs. Let us denote the DIACs by  $\omega_P^*$  and  $\omega_M^*$ , respectively. Then, we must show that

$$s^2\omega_P^* = \omega_M^*. \tag{116}$$

If we denote the elements of  $\omega_M^*$  by  $\omega_{ij}$  and define  $d' = 1/d$ , (116) results in the following six equations:

$$\omega_{11} = s^2a_x^2(1 - \cos^2 \rho \sin^2 \tau) + 2s^2a_xc_xd' \sin \rho \sin \tau + s^2c_x^2d'^2 \sin^2 \tau \tag{117}$$

$$\omega_{22} = s^2a_y^2(1 - \sin^2 \rho \sin^2 \tau) + 2s^2a_yc_yd' \cos \rho \sin \tau + s^2c_y^2d'^2 \sin^2 \tau \tag{118}$$

$$\omega_{33} = s^2d'^2 \sin^2 \tau \tag{119}$$

$$\omega_{12} = -s^2a_xa_y \cos \rho \sin \rho \sin^2 \tau + s^2a_xc_yd' \sin \rho \sin \tau - s^2a_yc_xd' \cos \rho \sin \tau + s^2c_xc_yd'^2 \sin^2 \tau \tag{120}$$

$$\omega_{13} = s^2a_xd' \sin \rho \sin \tau + s^2c_xd'^2 \sin^2 \tau \tag{121}$$

$$\omega_{23} = -s^2a_yd' \cos \rho \sin \tau + s^2c_yd'^2 \sin^2 \tau, \tag{122}$$

where  $\cos^2 \theta$  and  $\sin^2 \theta$  are abbreviations for  $(\cos \theta)^2$  and  $(\sin \theta)^2$ , respectively.

We have six equations containing eight unknown parameters. Until now, we have not distinguished between  $S_M$  and  $S_M^*$ .

We will now derive the finite set of maps  $\mathcal{A}$  on the set  $S_M^*$ . To do so, we will need to fix some parameters in order to decompose the matrix uniquely. We choose to fix the location of the principal point  $(c_x, c_y)^T$ , i.e., our set  $\mathcal{A}$  will effectively be parameterized by  $(c_x, c_y)^T$ :

$$\phi_{(c_x, c_y)_i} : U_i \rightarrow \mathbb{R}^n, \quad i \in I, \tag{123}$$

where  $I$  is a finite set of indices.

By using (117)–(122), we can describe the set  $D$  of measure 0 that we excluded from  $S_M$  in the parameter space. We have

$$\omega_{12}\omega_{33} - \omega_{13}\omega_{23} = -s^4a_xa_yd'^2 \cos \rho \sin \rho \sin^2 \tau \times (\sin^2 \tau - 1) \tag{124}$$

$$\omega_{11}\omega_{33} - \omega_{13}^2 = -s^4a_x^2d'^2 \sin^2 \tau \times (\cos^2 \rho \sin^2 \tau + \sin^2 \rho - 1) \tag{125}$$

$$\omega_{22}\omega_{33} - \omega_{23}^2 = -s^4a_y^2d'^2 \sin^2 \tau \times (\sin^2 \rho \sin^2 \tau + \cos^2 \rho - 1). \tag{126}$$

Thus,  $D$  corresponds to the set in the parameter space where  $\sin \tau = 0, \sin \tau = 1, \sin \rho = 0$ , or  $\cos \rho = 0$ . Therefore, we assume that  $\sin \tau \neq 0, \sin \tau \neq 1, \sin \rho \neq 0$ , and  $\cos \rho \neq 0$ , i.e.,  $\tau \neq k\pi, \tau \neq (4k + 1)\pi/2$ , and  $\rho \neq k\pi/2, k \in \mathbb{Z}$ .

Furthermore, we assume that  $\mathbf{M}$  is indeed a general camera at infinity. In particular, we assume that  $\mathbf{M}$  does not represent an affine camera since Theorem 1 already proves that every affine camera can be interpreted as a bilateral telecentric tilt camera. This implies that  $\omega_{33} > 0$  and we can substitute (119) into (121) and (122) and solve for  $\sin \rho$  and  $\cos \rho$ , respectively:

$$\sin \rho = \frac{\omega_{13} - c_x\omega_{33}}{sa_x\sqrt{\omega_{33}}} \tag{127}$$

$$\cos \rho = \frac{c_y\omega_{33} - \omega_{23}}{sa_y\sqrt{\omega_{33}}}. \tag{128}$$

We can now substitute (127), (128), and (119) into (120) and solve for  $\sin^2 \tau$ :

$$\sin^2 \tau = \frac{\omega_{12}\omega_{33} - (c_x\omega_{23} + c_y\omega_{13})\omega_{33} + c_xc_y\omega_{33}^2}{\omega_{13}\omega_{23} - (c_x\omega_{23} + c_y\omega_{13})\omega_{33} + c_xc_y\omega_{33}^2}. \tag{129}$$

Obviously,  $0 \leq \sin^2 \tau \leq 1$  must hold to be able to solve (129) for  $\tau$ . Actually, we need to guarantee  $0 < \sin^2 \tau < 1$  since we excluded the cases  $\sin \tau = 0$  and  $\sin \tau = 1$ . Equation (129) only depends on  $c_x$  and  $c_y$  (besides  $\mathbf{M}$ ), which we can now choose such that  $0 < \sin^2 \tau < 1$  holds.

Note that in (129), both in the numerator and in the denominator the same terms are added to  $\omega_{12}\omega_{33}$  and  $\omega_{13}\omega_{23}$ , respectively. To find suitable values for  $c_x$  and  $c_y$ , we will consider different cases. The easiest case is  $0 < \frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} < 1$ . In this case, we can set  $(c_x, c_y)^T = (0, 0)^T$ .

One special case that we must examine is  $\omega_{12} = 0$ . Then,  $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} = 0$ , but since we assume  $\sin \tau \neq 0$ , we cannot set  $(c_x, c_y)^T = (0, 0)^T$ . Because  $\mathbf{M} \notin D$ , (105) does not hold, which implies in this case that  $\omega_{13} \neq 0$  and  $\omega_{23} \neq 0$ . Then, without loss of generality, we can set  $c_y = 0$  and choose  $c_x < 0$  arbitrarily if  $\omega_{13} > 0$  and  $c_x > 0$  arbitrarily if  $\omega_{13} < 0$ . Alternatively, we could set  $c_x = 0$  and choose  $c_y < 0$  if  $\omega_{23} > 0$  and  $c_y > 0$  if  $\omega_{23} < 0$ .

Now, assume that  $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} \notin [0, 1]$ . We will derive inequality conditions for  $c_x$  and  $c_y$  that are of the form  $|c_{x,y}| > |\omega_{ij}/\omega_{kl}|$ . Because we want to show that one can choose  $(c_x, c_y)$  from a finite set of possible values in order to obtain  $\sin^2 \tau$  between 0 and 1 from (129), we want to prevent  $|c_x|$  and  $|c_y|$  from going to infinity, i.e., to prevent the relevant  $\omega_{ij}$  in the denominator from lying arbitrarily close to 0. Thus, instead of differentiating between  $\omega_{ij} = 0$  and  $\omega_{ij} \neq 0$  we will distinguish the cases  $|\omega_{ij}| > \epsilon$  and  $|\omega_{ij}| \leq \epsilon$  for some arbitrary but fixed  $\epsilon > 0$ . Here, we can choose  $\epsilon > 0$  small enough such that  $\omega_{33} > \epsilon$ . Such an  $\epsilon$  exists since we assume  $\omega_{33} > 0$ , as already mentioned above.

Let us examine the case  $|\omega_{23}| > \epsilon$  in detail. We can set  $c_y = 0$ . For  $c_x$ , we distinguish the following cases (keep in mind that  $\omega_{33} > 0$ ):

- $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} < 0$ 
  - $\omega_{12} < 0$ . Then,  $\omega_{12}\omega_{33} < 0 < \omega_{13}\omega_{23}$ . We want to choose  $c_x$  such that the numerator of  $\sin^2 \tau$  is positive. Then, also the denominator is positive and still greater than the numerator and  $0 < \sin^2 \tau < 1$ . Thus, we must choose:

$$c_x \begin{cases} \leq -3/\epsilon < \omega_{12}/\omega_{23} & \text{if } \omega_{23} > 0 \\ \geq 3/\epsilon > \omega_{12}/\omega_{23} & \text{if } \omega_{23} < 0 \end{cases} \quad (130)$$

- $\omega_{12} > 0$ . Then,  $\omega_{12}\omega_{33} > 0 > \omega_{13}\omega_{23}$ . We want to choose  $c_x$  such that the numerator of  $\sin^2 \tau$  is negative. Then, also the denominator is negative and still smaller than the numerator and  $0 < \sin^2 \tau < 1$ . Thus, we must choose:

$$c_x \begin{cases} \geq 3/\epsilon > \omega_{12}/\omega_{23} & \text{if } \omega_{23} > 0 \\ \leq -3/\epsilon < \omega_{12}/\omega_{23} & \text{if } \omega_{23} < 0 \end{cases} \quad (131)$$

- $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} > 1$ 
  - $\omega_{12} < 0$ . Note that, since  $\omega_{12}\omega_{33} < 0$ , also  $\omega_{13}\omega_{23}$  must be  $< 0$  for their ratio to become positive. Then,  $\omega_{12}\omega_{33} < \omega_{13}\omega_{23} < 0$ . We want to choose  $c_x$  such that the numerator of  $\sin^2 \tau$  is positive. Then, also the denominator is positive and still greater than the numerator and  $0 < \sin^2 \tau < 1$ . Thus, we must choose:

$$c_x \begin{cases} \leq -3/\epsilon < \omega_{12}/\omega_{23} & \text{if } \omega_{23} > 0 \\ \geq 3/\epsilon > \omega_{12}/\omega_{23} & \text{if } \omega_{23} < 0 \end{cases} \quad (132)$$

- $\omega_{12} > 0$ . Since  $\omega_{12}\omega_{33} > 0$ , also  $\omega_{13}\omega_{23}$  must be  $> 0$  for their ratio to become positive. Then,  $\omega_{12}\omega_{33} > \omega_{13}\omega_{23} > 0$ . We want to choose  $c_x$  such that the numerator of  $\sin^2 \tau$  is negative. Then, also the denominator is negative and still smaller than the numerator and  $0 < \sin^2 \tau < 1$ . Thus, we must choose:

$$c_x \begin{cases} \geq 3/\epsilon > \omega_{12}/\omega_{23} & \text{if } \omega_{23} > 0 \\ \leq -3/\epsilon < \omega_{12}/\omega_{23} & \text{if } \omega_{23} < 0 \end{cases} \quad (133)$$

In each case, the second inequality follows from  $|\omega_{12}| \leq 3$  and  $|\omega_{23}| > \epsilon$ .

Summarized, if  $|\omega_{23}| > \epsilon$ , we can choose:

$$c_x = k \cdot \frac{3}{\epsilon} \cdot \text{sgn}(\omega_{12}\omega_{23}) \quad (134)$$

$$c_y = 0 \quad (135)$$

for some  $k \geq 1$ .

Now let us consider the case  $|\omega_{23}| \leq \epsilon$ . We see that if  $|\omega_{13}| > \epsilon$ , we can switch the roles of  $c_x$  and  $c_y$  and  $\omega_{13}$  and  $\omega_{23}$  and obtain:

$$c_x = 0 \quad (136)$$

$$c_y = k \cdot \frac{3}{\epsilon} \cdot \text{sgn}(\omega_{12}\omega_{13}) \quad (137)$$

for some  $k \geq 1$ .

We can summarize the choice of  $(c_x, c_y)^T$  for some  $k \geq 1$  as follows:

$\omega_{33} > \epsilon$	$c_x$	$c_y$
$\omega_{12} \neq 0,  \omega_{23}  > \epsilon$	$k \cdot 3/\epsilon \cdot \text{sgn}(\omega_{12}\omega_{23})$	0
$\omega_{12} \neq 0,  \omega_{13}  > \epsilon$	0	$k \cdot 3/\epsilon \cdot \text{sgn}(\omega_{12}\omega_{13})$
$\omega_{12} = 0,  \omega_{13}  > \epsilon$	$-k \cdot 3/\epsilon \cdot \text{sgn}(\omega_{13})$	0
$\omega_{12} = 0,  \omega_{23}  > \epsilon$	0	$-k \cdot 3/\epsilon \cdot \text{sgn}(\omega_{23})$

Note that the last line is redundant and is included only for reasons of symmetry. As discussed above,  $\omega_{12} = 0$  implies that both  $\omega_{13}$  and  $\omega_{23}$  are  $\neq 0$  and hence their absolute values are  $> \epsilon$  (under the assumption that  $\sin \tau \neq 0$ ).

If now both  $|\omega_{23}| \leq \epsilon$  and  $|\omega_{13}| \leq \epsilon$ , we must choose  $c_x \neq 0$  and  $c_y \neq 0$  to obtain  $\sin^2 \tau \in [0, 1]$ . Note that we still assume  $\omega_{33} > \epsilon$ . We will only derive a valid choice for  $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} < 0$  in detail here. Again, we distinguish different cases:

- $\omega_{12} < 0, \omega_{23} \geq 0$ . We want to choose  $c_x$  and  $c_y$  such that the numerator of  $\sin^2 \tau$  is positive. Then, also the denominator is positive and still greater than the numerator and  $0 < \sin^2 \tau < 1$ . Thus, we start with setting  $c_x = c/\epsilon$  with  $c > \epsilon$ . With this, we can write the numerator of (129) as

$$\begin{aligned} & \omega_{12}\omega_{33} - \frac{c}{\epsilon}\omega_{23}\omega_{33} - c_y\omega_{13}\omega_{33} + c_y\frac{c}{\epsilon}\omega_{33}^2 \\ & \geq \omega_{12}\omega_{33} - c\omega_{33} - c_y\omega_{13}\omega_{33} + c_y\frac{c}{\epsilon}\omega_{33}^2 \quad (138) \\ & \stackrel{!}{>} 0, \end{aligned}$$

which can be reformulated as

$$c_y > \frac{c - \omega_{12}}{\frac{c}{\epsilon}\omega_{33} - \omega_{13}} \quad (139)$$

because  $\frac{c}{\epsilon}\omega_{33} - \omega_{13} \geq 0$ . Using the fact that  $|\omega_{ij}| \leq 3$ , we have  $(c - \omega_{12})/(\frac{c}{\epsilon}\omega_{33} - \omega_{13}) \leq (c + 3)/(c - \epsilon)$  and we can choose

$$c_y > \frac{c + 3}{c - \epsilon} \quad (140)$$

–  $\omega_{12} < 0, \omega_{23} < 0$ . Again, we want to choose  $c_x$  and  $c_y$  such that the numerator of  $\sin^2 \tau$  is positive. To be able to do a similar estimate as in (138), we set  $c_x = -c/\epsilon$  with  $c > \epsilon$ . Then, we have

$$\begin{aligned} &\omega_{12}\omega_{33} + \frac{c}{\epsilon}\omega_{23}\omega_{33} - c_y\omega_{13}\omega_{33} - c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\geq \omega_{12}\omega_{33} - c\omega_{33} - c_y\omega_{13}\omega_{33} - c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\stackrel{!}{>} 0, \end{aligned} \tag{141}$$

which can be reformulated as

$$c_y < \frac{c - \omega_{12}}{-\frac{c}{\epsilon}\omega_{33} - \omega_{13}} \tag{142}$$

because  $-\frac{c}{\epsilon}\omega_{33} - \omega_{13} \leq -c + \epsilon < 0$ . Furthermore, we have  $(c - \omega_{12})/(-\frac{c}{\epsilon}\omega_{33} - \omega_{13}) \geq (c + 3)/(\epsilon - c)$  and we can choose

$$c_y < \frac{c + 3}{\epsilon - c}. \tag{143}$$

–  $\omega_{12} > 0, \omega_{23} \geq 0$ . Here, we want the numerator of  $\sin^2 \tau$  to be negative. We set  $c_x = -c/\epsilon$  with  $c > \epsilon$ . Then, we have

$$\begin{aligned} &\omega_{12}\omega_{33} + \frac{c}{\epsilon}\omega_{23}\omega_{33} - c_y\omega_{13}\omega_{33} - c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\leq \omega_{12}\omega_{33} + c\omega_{33} - c_y\omega_{13}\omega_{33} - c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\stackrel{!}{<} 0, \end{aligned} \tag{144}$$

which can be reformulated as

$$c_y > \frac{-c - \omega_{12}}{-\frac{c}{\epsilon}\omega_{33} - \omega_{13}} \tag{145}$$

because  $-\frac{c}{\epsilon}\omega_{33} - \omega_{13} \leq -c + \epsilon < 0$ . Furthermore, we have  $(-c - \omega_{12})/(-\frac{c}{\epsilon}\omega_{33} - \omega_{13}) \leq (c + 3)/(c - \epsilon)$  and we can choose

$$c_y > \frac{c + 3}{c - \epsilon}. \tag{146}$$

–  $\omega_{12} > 0, \omega_{23} < 0$ . Again, we want the numerator of  $\sin^2 \tau$  to be negative. We set  $c_x = c/\epsilon$  with  $c > \epsilon$ . Then, we have

$$\begin{aligned} &\omega_{12}\omega_{33} - \frac{c}{\epsilon}\omega_{23}\omega_{33} - c_y\omega_{13}\omega_{33} + c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\leq \omega_{12}\omega_{33} + c\omega_{33} - c_y\omega_{13}\omega_{33} + c_y\frac{c}{\epsilon}\omega_{33}^2 \\ &\stackrel{!}{<} 0, \end{aligned} \tag{147}$$

which can be reformulated as

$$c_y < \frac{-c - \omega_{12}}{\frac{c}{\epsilon}\omega_{33} - \omega_{13}} \tag{148}$$

because  $\frac{c}{\epsilon}\omega_{33} - \omega_{13} \geq c - \epsilon > 0$ . Furthermore, we have  $(-c - \omega_{12})/(\frac{c}{\epsilon}\omega_{33} - \omega_{13}) \geq (c + 3)/(\epsilon - c)$  and we can choose

$$c_y < \frac{c + 3}{\epsilon - c}. \tag{149}$$

For the case  $\frac{\omega_{12}\omega_{33}}{\omega_{13}\omega_{23}} > 1$ , we obtain exactly the same valid choices for  $c_x$  and  $c_y$  depending on the signs of  $\omega_{12}$  and  $\omega_{23}$ . Thus, for some  $c > \epsilon$  and  $k > 1$ , we can summarize:

		$c_x$	$c_y$
$ \omega_{23}  \leq \epsilon,$ $ \omega_{13}  \leq \epsilon,$ $\omega_{33} > \epsilon$			
$\omega_{12} > 0$	$\omega_{23} \geq 0$	$-c/\epsilon$	$k \cdot (c + 3)/(c - \epsilon)$
	$\omega_{23} < 0$	$c/\epsilon$	$k \cdot (c + 3)/(\epsilon - c)$
$\omega_{12} < 0$	$\omega_{23} \geq 0$	$c/\epsilon$	$k \cdot (c + 3)/(c - \epsilon)$
	$\omega_{23} < 0$	$-c/\epsilon$	$k \cdot (c + 3)/(\epsilon - c)$

Now, it only remains to examine the case  $|\omega_{23}| \leq \epsilon, |\omega_{13}| \leq \epsilon,$  and  $|\omega_{33}| \leq \epsilon$ . This means that the entries of the last row of  $\mathbf{M}_3$ , tend to 0 since we can choose  $\epsilon$  arbitrarily small. The matrix  $\mathbf{M}$  thus is not a general camera at infinity anymore, but an affine camera and we can apply the decomposition result for affine cameras from Theorem 1.

Consequently, we can conclude that, if  $\mathbf{M}$  indeed represents a general camera at infinity and not an affine camera, the principal point  $(c_x, c_y)^T$  can be chosen from a finite set of possible values. These values are not unique since we have two degrees of freedom for  $c$  and  $k$ . But we can fix these parameters and thus obtain the foundation for the index set of  $\mathcal{A}$  on  $S_M^*$  that maps the matrix  $\mathbf{M}$  to the parameters of the object-side telecentric tilt camera.

Setting the values for  $c_x$  and  $c_y$  according to the computations above for which (129) fulfills the condition  $\sin^2 \tau \in (0, 1)$ , we can solve (129) for  $\tau$ . We have

$$\tau = \arcsin \sqrt{\frac{\omega_{12}\omega_{33} - (c_x\omega_{23} + c_y\omega_{13})\omega_{33} + c_x c_y \omega_{33}^2}{\omega_{13}\omega_{23} - (c_x\omega_{23} + c_y\omega_{13})\omega_{33} + c_x c_y \omega_{33}^2}} \tag{150}$$

and we can continue deriving the remaining unknown parameters. We can solve (121) and (122) for  $sa_x$  and  $sa_y$ , respectively:

$$sa_x = \frac{\omega_{13} - c_x \omega_{33}}{\sin \rho \sqrt{\omega_{33}}}, \tag{151}$$

$$sa_y = \frac{c_y \omega_{33} - \omega_{23}}{\cos \rho \sqrt{\omega_{33}}}, \tag{152}$$

and substitute (151) and (152) into (117) and (118), respectively, which we can both solve for  $\cos^2 \rho$ :

$$\cos^2 \rho = \frac{\omega_{11} \omega_{33} - \omega_{13}^2}{c_x^2 c_\tau^2 \omega_{33}^2 + (\omega_{11} - 2c_x c_\tau^2 \omega_{13}) \omega_{33} - s_\tau^2 \omega_{13}^2} \tag{153}$$

$$\cos^2 \rho = \frac{c_\tau^2 (\omega_{23}^2 - 2c_y \omega_{23} \omega_{33} + c_y^2 \omega_{33}^2)}{c_y^2 c_\tau^2 \omega_{33}^2 + (\omega_{22} - 2c_y c_\tau^2 \omega_{23}) \omega_{33} - s_\tau^2 \omega_{23}^2}, \tag{154}$$

where  $c_\tau^2 = \cos^2 \tau$  and  $s_\tau^2 = \sin^2 \tau$ .

Therefore, we have two possible solutions for  $\cos^2 \rho$ . To show that both equations lead to the same value of  $\cos^2 \rho$ , we must show that the right hand sides of (153) and (154) are identical. In fact, this is a consequence of  $\mathbf{M}_3$  being singular. To see this, one can expand the DIAC elements  $\omega_{ij}$  in terms of the matrix elements  $m_{ij}$ , express one of them, e.g.,  $m_{33}$ , by the others by means of  $\det \mathbf{M}_3 = 0$ , and substitute that together with (150) into (153) and (154). Since this is very tedious, it is best done using a computer algebra system. This shows that (153) and (154) are identical.

Furthermore, it must hold that  $0 \leq \cos^2 \rho \leq 1$  to be able to solve (153) for  $\rho$ . We can reformulate (153) as follows:

$$\cos^2 \rho = \frac{\omega_{11} \omega_{33} - \omega_{13}^2}{\cos^2 \tau (c_x \omega_{33} - \omega_{13})^2 + \omega_{11} \omega_{33} - \omega_{13}^2}. \tag{155}$$

Using the Cauchy–Schwarz inequality, which implies  $\omega_{13}^2 \leq \omega_{11} \omega_{33}$ , we can see directly from (155) that  $0 \leq \cos^2 \rho \leq 1$  holds. Indeed, it even holds that  $0 < \cos^2 \rho < 1$  since we know that (106) does not hold and therefore  $\omega_{13}^2$  is strictly smaller than  $\omega_{11} \omega_{33}$ . We then have

$$\rho = \arccos \sqrt{\frac{\omega_{11} \omega_{33} - \omega_{13}^2}{\cos^2 \tau (c_x \omega_{33} - \omega_{13})^2 + \omega_{11} \omega_{33} - \omega_{13}^2}}. \tag{156}$$

However, this returns  $\rho$  only between 0 and  $\pi/2$ . To determine  $\rho$  between 0 and  $2\pi$ , we can examine the signs of (121) and (122). This shows that

- $0 < \rho < \pi/2$  if  $\omega_{13} - c_x \omega_{33} \geq 0$  and  $\omega_{23} - c_y \omega_{33} \leq 0$ ,
- $\pi/2 < \rho < \pi$  if  $\omega_{13} - c_x \omega_{33} \geq 0$  and  $\omega_{23} - c_y \omega_{33} \geq 0$ ,
- $\pi < \rho < 3\pi/2$  if  $\omega_{13} - c_x \omega_{33} \leq 0$  and  $\omega_{23} - c_y \omega_{33} \geq 0$ ,
- $3\pi/2 < \rho < 2\pi$  if  $\omega_{13} - c_x \omega_{33} \leq 0$  and  $\omega_{23} - c_y \omega_{33} \leq 0$ .

Note that we cannot determine  $s$  from (117)–(122). We can therefore initially set  $s = 1$  and solve (151) and (152) for  $a_x$  and  $a_y$  as well as (119) for  $d'$ . We can now compute the tilt homography  $\mathbf{H}$  and the calibration matrix  $\mathbf{K}$ . Conceptually, we obtain

$$\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} & 0 \\ h_{21} & h_{22} & 0 \\ sh_{31} & sh_{32} & h_{33} \end{pmatrix} \tag{157}$$

and

$$\mathbf{K} = \begin{pmatrix} sa_x & 0 & c_x \\ 0 & sa_y & c_y \\ 0 & 0 & 1 \end{pmatrix}, \tag{158}$$

where  $s$  is the true scaling factor. By multiplying  $\mathbf{P}_{op}$  and  $\mathbf{M}$  from the left by  $\mathbf{H}^{-1} \mathbf{K}^{-1}$ , we obtain

$$\mathbf{H}^{-1} \mathbf{K}^{-1} \mathbf{M} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & s \end{pmatrix}. \tag{159}$$

Consequently, the scale factor  $s$  can be obtained from the element (3, 4) of this matrix. The true values of  $d'$ ,  $a_x$ , and  $a_y$  can be obtained by dividing the preliminary values derived above by  $s$ . The missing third row of the rotation matrix can be reconstructed by the vector product of the first two rows. Since the orthographic projection in object space (8), and therefore also (110)–(115), do not depend on  $t_z$ , we can set  $t_z$  to any value, e.g., to 0.

This gives us the set of maps  $\mathcal{A} = \{\phi_i\}_{i \in I}$  that is defined on  $S_M^*$ . We have seen that with  $\mathcal{A}$  any matrix  $\mathbf{M} \in S_M^*$  can be decomposed into the parameters of an object-side telecentric tilt camera by giving an explicit computation of the parameters.  $\square$

**Lemma 2** Any matrix  $\mathbf{M} \in D = S_M \setminus S_M^*$  can be regarded as an object-side telecentric tilt camera, i.e., there exist mappings that map  $\mathbf{M}$  to the parameter space of object-side telecentric tilt cameras. In contrast to Lemma 1, these mappings are not unique.

*Proof* We now consider the set  $D$  of matrices that we have excluded from  $S_M$  in Lemma 1.

In the proof of Lemma 1, we have shown that the set  $D$  is described by  $\sin \tau = 0$ ,  $\sin \tau = 1$ ,  $\sin \rho = 0$ , or  $\cos \rho = 0$  in the parameter space.

Let us start by examining the case  $\sin \tau = 0$ , i.e.,  $\tau = k\pi, k \in \mathbb{Z}$ . It is easy to see that  $\sin \tau = 0$  is equivalent to  $\omega_{12} = \omega_{13} = \omega_{23} = \omega_{33} = 0$ . Consequently, the camera is an affine camera, which we have excluded from the theorem. Affine cameras can be decomposed using the decomposition in the proof of Theorem 1.



Next, assuming  $\sin \tau \neq 0$ , we consider the cases  $\sin \tau = 1$ ,  $\sin \rho = 0$ , or  $\cos \rho = 0$ . We have already seen that the following equations hold:

$$\omega_{12}\omega_{33} - \omega_{13}\omega_{23} = -s^4 a_x a_y d'^2 \cos \rho \sin \rho \sin^2 \tau \times (\sin^2 \tau - 1) \tag{160}$$

$$\omega_{11}\omega_{33} - \omega_{13}^2 = -s^4 a_x^2 d'^2 \sin^2 \tau \times (\cos^2 \rho \sin^2 \tau + \sin^2 \rho - 1) \tag{161}$$

$$\omega_{22}\omega_{33} - \omega_{23}^2 = -s^4 a_y^2 d'^2 \sin^2 \tau \times (\sin^2 \rho \sin^2 \tau + \cos^2 \rho - 1). \tag{162}$$

Because we assume  $s > 0$ ,  $a_x > 0$ ,  $a_y > 0$ , and  $d' > 0$ , we have:

$$\begin{aligned} -\omega_{12}\omega_{33} - \omega_{13}\omega_{23} = 0 &\Leftrightarrow \sin \tau = 1, \cos \rho = 0, \text{ or } \sin \rho = 0, \\ -\omega_{11}\omega_{33} - \omega_{13}^2 = 0 &\Leftrightarrow \sin \tau = 1 \text{ or } \cos \rho = 0, \\ -\omega_{22}\omega_{33} - \omega_{23}^2 = 0 &\Leftrightarrow \sin \tau = 1 \text{ or } \sin \rho = 0. \end{aligned}$$

Therefore, if (105)–(107) hold, we are in the case  $\sin \tau = 1$  and can, for example, ignore (117), (118), and (120) since they are redundant to (119), (121), and (122). By setting  $s = 1$ , we can solve (119) for  $d'$ . Then, we are left with (121) and (122). We can, for example, set  $c_x = 0$ ,  $c_y = 0$ ,  $\rho = k\pi/4$  for  $k \in \{1, 3, 5, 7\}$ , depending on the signs of  $\omega_{13}$  and  $\omega_{23}$ , and then solve (121) and (122) for  $a_x$  and  $a_y$ . We can then proceed as in the proof of Lemma 1 to compute (159) and determine the pose parameters and the true values of  $s$ ,  $d'$ ,  $a_x$ , and  $a_y$ .

Furthermore, if (105) and either (106) or (107), but not both, hold, we have the case that  $\cos \rho = 0$  or  $\sin \rho = 0$ , respectively, and  $\sin \tau \neq 1$ . We will only describe the case  $\cos \rho = 0$  in detail here. The case  $\sin \rho = 0$  can be handled analogously.

If  $\cos \rho = 0$ , (122) reduces to

$$\omega_{23} = s^2 c_y d'^2 \sin^2 \tau. \tag{163}$$

Since we have an explicit equation for  $c_y$  here, we cannot apply the approach from the proof of Lemma 1, where we gave a finite set of possible values for  $c_x$  and  $c_y$ .

By substituting  $s^2 d'^2 \sin^2 \tau$  with  $\omega_{33}$ , we can solve (163) directly for  $c_y$ . To obtain a unique solution, we can require  $c_x = 0$ , which reduces (117) and (121) to

$$\omega_{11} = s^2 a_x^2 \tag{164}$$

$$\omega_{13} = s^2 a_x d' \sin \rho \sin \tau. \tag{165}$$

We can solve (164) for  $a_x$  by setting  $s = 1$ . From the sign of  $\omega_{13}$ , we can determine  $\rho$ :  $\rho = \pi/2$  if  $\omega_{13} > 0$  and  $\rho = 3\pi/2$  if  $\omega_{13} < 0$ .

Rearranging (118) leads to

$$\cos^2 \tau = \frac{\omega_{22} - c_y^2 \omega_{33}}{s^2 a_y^2}. \tag{166}$$

Thus,  $\tau$  and  $a_y$  cannot be determined simultaneously (which was to be expected because of Proposition 5). To solve this degeneracy, we can choose  $a_y$  such that  $\cos^2 \tau \in [0, 1]$  since we have one degree of freedom left. We can, for example, set  $a_y = k\sqrt{\omega_{22} - c_y^2 \omega_{33}}$  for  $k > 1$ . This can be done since, because of (163) and the Cauchy–Schwarz inequality, the term in the square root is always nonnegative and even positive because we know that (107) does not hold. Then  $\tau$  only depends on  $k$ . We can solve (166) for  $\tau$  and (119) for  $d'$  and then proceed as usual to compute (159) and determine the true values of  $s$ ,  $d'$ ,  $a_x$ , and  $a_y$ .

Thus, in contrast to the computation of the parameters in the proof of Theorem 3, we must fix three instead of two parameters here, namely  $c_x$  (for  $\cos \rho = 0$ ) or  $c_y$  (for  $\sin \rho = 0$ ),  $\rho$ , and  $k \hat{=} \tau$ , to get a set of maps that maps  $D$  uniquely to the parameter space.

These decompositions of  $\mathbf{M} \in D$  form the mappings from  $D$  to the parameter space of object-side telecentric tilt cameras.  $\square$

*Remark 23* There are degeneracies in the set  $D = S_M \setminus S_M^*$  that do not allow the unique decomposition into the minimal parameterization. This is a kind of singularity and therefore not the whole set  $S_M$  can be described by a continuation from  $S_M^*$  to  $S_M$  of the finite set of maps  $\mathcal{A} = \{\phi_i\}_{i \in I}$  that was derived in the proof of Lemma 1.

*Remark 24* The proof of Lemma 1 provides an algorithm to decompose a general camera at infinity into the parameters of an object-side telecentric tilt camera using the minimum number of 10 parameters using a finite set of maps. As a consequence, the camera parameters of the decomposition may not correspond to the geometric intuition one might have about a particular problem. If it is desired to decompose a general camera at infinity into geometrically intuitive parameters, the principal point must be known, at least approximately. Then, the general camera at infinity can be normalized to this principal point by multiplying it from the left with a suitable calibration matrix that moves the principal point to  $(0, 0)^T$ . This allows the decomposition with  $(c_x, c_y)^T = (0, 0)^T$  in the proof of Lemma 1 to be used.

## References

Aggarwal, M., & Ahuja, N. (2002). A pupil-centric model of image formation. *International Journal of Computer Vision*, 48(3), 195–214.

- Ahn, S. J., Warnecke, H. J., & Kotowski, R. (1999). Systematic geometric image measurement errors of circular object targets: Mathematical formulation and correction. *Photogrammetric Record*, 16(93), 485–502.
- Albers, O., Poesch, A., & Reithmeier, E. (2015). Flexible calibration and measurement strategy for a multi-sensor fringe projection unit. *Optics Express*, 23(23), 29592–29607.
- Astarita, T. (2012). A Scheimpflug camera model for stereoscopic and tomographic PIV. In: *16th international symposium on applications of laser techniques to fluid mechanics*.
- Bauchau, O. A., & Trainelli, L. (2003). The vectorial parameterization of rotation. *Nonlinear Dynamics*, 32(1), 71–92.
- Beyerer, J., Puente León, F., & Frese, C. (2016). *Machine vision: Automated visual inspection: Theory, practice and applications*. Berlin: Springer.
- Blahusch, G., Eckstein, W., Steger, C., & Lanser, S. (1999). Algorithms and evaluation of a high precision tool measurement system. In: *5th international conference on quality control by artificial vision* (pp. 31–36).
- Brown, D. C. (1966). Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3), 444–462.
- Brown, D. C. (1971). Close-range camera calibration. *Photogrammetric Engineering*, 37(8), 855–866.
- Carpentier, J. (1901). Improvements in enlarging or like cameras. Great Britain Patent 1139.
- Claus, D., & Fitzgibbon, A. W. (2005). A rational function lens distortion model for general cameras. *IEEE Conference on Computer Vision and Pattern Recognition*, 1, 213–219.
- Cornic, P., Illoul, C., Sant, Y.L., Cheminet, A., Besnerais, G.L., & Champagnat, F. (2015). Calibration drift within a Tomo-PIV setup and self-calibration. In: *11th international symposium on particle image velocimetry*.
- Datta, A., Kim, J.S., & Kanade, T. (2009). Accurate camera calibration using iterative refinement of control points. In: *12th international conference on computer vision workshops* (pp. 1201–1208).
- DIN 1335:2003-12 (2003) Geometrische Optik—Bezeichnungen und Definitionen.
- Dold, J. (1996). Influence of large targets on the results of photogrammetric bundle adjustment. In *International archives of photogrammetry and remote sensing, vol XXXI, part B5* (pp. 119–123).
- Dold, J. (1997). Ein hybrides photogrammetrisches Industriemeßsystem höchster Genauigkeit und seine Überprüfung. Dissertation, Fakultät für Bauingenieur- und Vermessungswesen der Universität der Bundeswehr München, Neubiberg, Schriftenreihe Studiengang Vermessungswesen, Heft 54.
- Douxchamps, D., & Chihara, K. (2009). High-accuracy and robust localization of large control markers for geometric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), 376–383.
- Evens, L. (2008a). Depth of field for the tilted lens plane (online). [http://www.math.northwestern.edu/~len/photos/pages/tilt\\_dof.pdf](http://www.math.northwestern.edu/~len/photos/pages/tilt_dof.pdf). Accessed 10 June 2013.
- Evens, L. (2008b). View camera geometry (online). <http://www.math.northwestern.edu/~len/photos/pages/vc.pdf>. Accessed 10 June 2013.
- Fasogbon, P., Duviébourg, L., Lacaze, P. A., & Macaire, L. (2015). Intrinsic camera calibration equipped with Scheimpflug optical device. In F. Meriaudeau, O. Aubret (Eds.) *Twelfth international conference on quality control by artificial vision, Proceedings of SPIE 9534*
- Faugeras, O., & Luong, Q. T. (2001). *The geometry of multiple images: The laws that govern the formation of multiple images of a scene and some of their applications*. Cambridge: MIT Press.
- Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999). Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5), 476–480.
- Fitzgibbon, A. W. (2001). Simultaneous linear estimation of multiple view geometry and lens distortion. *IEEE Conference on Computer Vision and Pattern Recognition*, 1, 125–132.
- Fournel, T., Coudert, S., Fournier, C., & Ducotet, C. (2003). Stereoscopic particle image velocimetry using telecentric lenses. *Measurement Science and Technology*, 14(4), 494–499.
- Fournel, T., Lavest, J. M., Coudert, S., & Collange, F. (2004). Self-calibration of PIV video-cameras in Scheimpflug condition. In M. Stanislas, J. Westerweel, J. Kompenhans (Eds.) *Particle image velocimetry: Recent improvements* (pp. 391–405). Berlin: Springer.
- Fournel, T., Louhichi, H., Barat, C., & Menudet, J. F. (2006). Scheimpflug self-calibration based on tangency points. In *12th international symposium on flow visualization*.
- Gennery, D. B. (2001). Least-squares camera calibration including lens distortion and automatic editing of calibration points. In A. Gruen & T. S. Huang (Eds.), *Calibration and orientation of cameras in computer vision* (pp. 123–136). New York: Springer.
- Gennery, D. B. (2006). Generalized camera calibration including fish-eye lenses. *International Journal of Computer Vision*, 68(3), 239–266.
- Gennery, D. B., Litwin, T., Wilcox, B., Bon, B. (1987). Sensing and perception research for space telerobotics at JPL. In *International conference on robotics and automation* (pp. 311–317).
- Gerdas, R., Otterbach, R., & Kammüller, R. (1993a). Kalibrierung eines digitalen Bildverarbeitungssystems mit CCD-Kamera Teil 1: Modellbildung und Verfahren. *Technisches Messen*, 60(6), 255–261.
- Gerdas, R., Otterbach, R., & Kammüller, R. (1993b). Kalibrierung eines digitalen Bildverarbeitungssystems mit CCD-Kamera Teil II: Implementierung und Fehleranalyse. *Technisches Messen*, 60(7/8), 283–288.
- Gluckman, J., & Nayar, S. K. (2001). Rectifying transformations that minimize resampling effects. *IEEE Conference on Computer Vision and Pattern Recognition*, 1, 111–117.
- Gross, H. (2005). *Handbook of optical systems, volume 1: Fundamentals of technical optics*. Weinheim: Wiley-VCH.
- Haig, C., Heipke, C., & Wiggenhagen, M. (2006). Lens inclination due to instable fixings detected and verified with VDI/VDE 2634 part 1. In *International archives of photogrammetry, remote sensing and spatial information sciences* (Vol. XXXVI, part 5).
- Hamrouni, S., Louhichi, H., Aissia, H. B., & Elhajem, M. (2012). A new method for stereo-cameras self-calibration in Scheimpflug condition. In *15th international symposium on flow visualization*.
- Hanning, T. (2011). *High precision camera calibration*. Wiesbaden: Vieweg+Teubner Verlag.
- Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision* (2nd ed.). Cambridge: Cambridge University Press.
- Heikkilä, J. (1997). Accurate camera calibration and feature based 3-d reconstruction from monocular image sequences. Dissertation, Department of Electrical Engineering, University of Oulu, Acta Universitatis Oulensis Technica C 108.
- Heikkilä, J. (2000). Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10), 1066–1077.
- Konrath, R., & Schröder, W. (2000). Stereoscopic particle-image velocimetry (PIV): A new approach using telecentric lenses. In *10th international symposium on applications of laser techniques to fluid mechanics*.
- Konrath, R., & Schröder, W. (2002). Telecentric lenses for imaging in particle image velocimetry: A new stereoscopic approach. *Experiments in Fluids*, 33(5), 703–708.

- Kumar, A., & Ahuja, N. (2014a). Generalized pupil-centric imaging and analytical calibration for a non-frontal camera. In *IEEE conference on computer vision and pattern recognition* (pp. 3970–3977).
- Kumar, A., & Ahuja, N. (2014b). Generalized radial alignment constraint for camera calibration. In *22nd international conference on pattern recognition* (pp. 184–189).
- Kumar, A., & Ahuja, N. (2015). On the equivalence of moving entrance pupil and radial distortion for camera calibration. In *International conference on computer vision* (pp. 2345–2353).
- Lanser, S. (1997). Modellbasierte Lokalisation gestützt auf monokulare Videobilder. Dissertation, Forschungs- und Lehrereinheit Informatik IX, Technische Universität München.
- Lanser, S., Zierl, C., & Beuthhauser, R. (1995). Multibildkalibrierung einer CCD-Kamera. In G. Sagerer, S. Posch, & F. Kummert (Eds.), *Mustererkennung* (pp. 481–491). Springer, Informatik aktuell.
- Legarda, A., Izaguirre, A., Arana, N., & Iturrospe, A. (2011). A new method for Scheimpflug camera calibration. In *10th international workshop on electronics, control, measurement and signals*
- Legarda, A., Izaguirre, A., Arana, N., & Iturrospe, A. (2013). Comparison and error analysis of the standard pin-hole and Scheimpflug camera calibration models. In *11th international workshop of electronics, control, measurement, signals and their application to mechatronics*.
- Lenhardt, K. (2006). Optical systems in machine vision. In A. Hornberg (Ed.), *Handbook of machine vision*. Weinheim: Wiley-VCH.
- Lenz, R. (1987). Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit. In E. Paulus (Ed.), *Mustererkennung* (pp. 212–216). Springer, Informatik-Fachberichte.
- Lenz, R. (1988). Videometrie mit CCD-Sensoren und ihre Anwendung in der Robotik. Habilitationsschrift, Lehrstuhl für Nachrichtentechnik der Technischen Universität München.
- Lenz, R., & Fritsch, D. (1990). Accuracy of videometry with CCD sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 45(2), 90–110.
- Li, J., Guo, Y., Zhu, J., Lin, X., Xin, Y., Duan, K., et al. (2007). Large depth-of-view portable three-dimensional laser scanner and its segmental calibration for robot vision. *Optics and Lasers in Engineering*, 45(11), 1077–1087.
- Louhichi, H., Fournel, T., Lavest, J. M., & Aissia, H. B. (2006). Camera self-calibration in Scheimpflug condition for air flow investigation. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, & T. Malzbender (Eds.), *Advances in visual computing* (pp. 891–900). Berlin: Springer.
- Louhichi, H., Fournel, T., Lavest, J. M., & Aissia, H. B. (2007). Self-calibration of Scheimpflug cameras: an easy protocol. *Measurement Science and Technology*, 18(8), 2616–2622.
- Luhmann, T., Robson, S., Kyle, S., & Boehm, J. (2014). *Close-range photogrammetry and 3D imaging* (2nd ed.). Berlin: De Gruyter.
- Mallon, J., & Whelan, P. F. (2007). Which pattern? Biasing aspects of planar calibration patterns and detection methods. *Pattern Recognition Letters*, 28(9), 921–930.
- Merklinger, H. M. (2010). Focusing the view camera: A scientific way to focus the view camera and estimate depth of field. Version 1.6.1; online. <http://www.trenholm.org/hmmerk/FVC161.pdf>. Accessed 25 June 2013.
- Mikhail, E. M., Bethel, J. S., & McGlone, J. C. (2001). *Introduction to modern photogrammetry*. New York: Wiley.
- Morawiec, A., & Field, D. P. (1996). Rodrigues parameterization for orientation and misorientation distributions. *Philosophical Magazine A*, 73(4), 1113–1130.
- Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). New York: Springer.
- Otepka, J. (2004). Precision target mensuration in vision metrology. Dissertation, Fakultät für Mathematik und Geoinformation, Technische Universität Wien
- Otepka, J. O., & Fraser, C. S. (2004). Accuracy enhancement of vision metrology through automatic target plane determination. In *International archives of photogrammetry and remote sensing* (Vol. XXXV, part B5, pp. 873–879).
- Peng, J., Wang, M., Deng, D., Liu, X., Yin, Y., & Peng, X. (2015). Distortion correction for microscopic fringe projection system with Scheimpflug telecentric lens. *Applied Optics*, 54(34), 10055–10062.
- Prasad, A. K. (2000). Stereoscopic particle image velocimetry. *Experiments in Fluids*, 29(2), 103–116.
- Scheimpflug, T. (1902a). Einrichtung zur methodischen Verzerrung ebener Bilder auf optischem Wege mit beliebigen Linsensystemen. Austrian Patent 22923
- Scheimpflug, T. (1902b). Verfahren und Apparat zur methodischen Verzerrung ebener Bilder auf photographischem Wege mit beliebigen Objektiven. Austrian Patent 20937
- Scheimpflug, T. (1902c). Verfahren und Apparat zur methodischen Verzerrung ebener Bilder auf photographischem Wege mit beliebigen Objektiven. Austrian Patent 20299
- Scheimpflug, T. (1903a). Apparatus for the systematic alteration of plane pictures by means of lenses or mirrors. *United States Patent*, 752596
- Scheimpflug, T. (1903b). Einrichtung zur methodischen Verzerrung ebener Bilder auf optischem Wege mit beliebigen Linsensystemen. Austrian Patent 22924
- Scheimpflug, T. (1903c). Method of distorting plane images by means of lenses or mirrors. *United States Patent*, 751347
- Scheimpflug, T. (1904). Improved method and apparatus for the systematic alteration or distortion of plane pictures and images by means of lenses and mirrors for photography and other purposes. Great Britain Patent 1196
- Spong, M. W., Hutchinson, S., & Vidyasagar, M. (2006). *Robot modeling and control*. New York: Wiley.
- Steger, C. (1998). Unbiased extraction of curvilinear structures from 2d and 3d images. Dissertation, Fakultät für Informatik, Technische Universität München.
- Steger, C. (2000). Subpixel-precise extraction of lines and edges. In *International archives of photogrammetry and remote sensing* (Vol. XXXIII, part B3, pp. 141–156).
- Steger, C., Ulrich, M., & Wiedemann, C. (2008). *Machine vision algorithms and applications*. Weinheim: Wiley-VCH.
- Sturm, P., Ramalingam, S., Tardif, J. P., Gasparini, S., & Barreto, J. (2010). Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision*, 6(1–2), 1–183.
- Sturm, P. F., & Maybank, S. J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. *IEEE Conference on Computer Vision and Pattern Recognition*, 1, 432–437.
- Vo, M., Wang, Z., Luu, L., & Ma, J. (2011). Advanced geometric camera calibration for machine vision. *Optical Engineering*, 50(11), 1105031–1105033.
- Walker, S. (2002). Two-axes Scheimpflug focusing for particle image velocimetry. *Measurement Science and Technology*, 13(1), 1–12.
- Wang, J., Shi, F., Zhang, J., & Liu, Y. (2008). A new calibration model of camera lens distortion. *Pattern Recognition*, 41(2), 607–615.
- Wheeler, R. E. (2003). Notes on view camera geometry (online). <http://www.bobwheeler.com/photo/ViewCam.pdf>. Accessed 19 June 2013.
- Willert, C. (1997). Stereoscopic digital particle image velocimetry for application in wind tunnel flows. *Measurement Science and Technology*, 8(12), 1465–1479.