

## Chapter 1

### BACKGROUND

*To suppose that the eye, with all its inimitable contrivances ... could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree.*

— Charles Darwin (Origin of Species)

The odds of two randomly selected irises, the donut-shaped colored tissue of the eye, to exactly match is 1 in  $10^{72}$  [1]. This staggering *singularity* of the iris pattern arises from the tumultuous and chaotic morphogenesis of its minutiae during the fetal development. Once formed, the *externally visible* iris pattern remains *stable* over a long time, safely *protected* under the transparent cornea. These unique features of the human iris make it an ideal biometric (biological measurement) technology [2]. Indeed, iris recognition, which works by matching the structural pattern of the iris, is a powerful surveillance technology that has gained unprecedented success in a multitude of personal identification and access authentication applications [3,4].

However, the use of iris recognition technology has been mostly restricted to fully-cooperative and well-controlled environments. Subjects are often expected to remain stationary at a fixed *standoff* distance or move slowly within a restricted path (e.g. through a portal) during the iris acquisition phase. These constraints impede the deployment of iris recognition in large scale, unrestricted environments such as crowded stadiums, public transits, border security, etc. The failure to recognize irises outside a small *capture volume* is still an unsolved problem [5–7].

The demand for automatic surveillance systems that can provide security day-after-day is eminent in the disorderly world we inhabit today. Iris recognition is promising not only because it is an ideal biometric but also because it does not require physical contact with the subject, and it does not entail human intervention. However, the central impediment to deploying iris biometric monitoring in large-scale, unrestricted environments is its limited *capture volume* [5–7].

Assuming the acquisition optics has sufficient magnification (discussed in [Sec. 1.4](#) and illustrated in [Figure 1.10](#)), the capture volume is restricted by the depth of field (DOF) of the optics. A lens can focus only on a single surface in the object space. This surface is often a plane—the plane of sharp focus—due to the common planar surface structure of sensors. *The DOF is defined as the region (range of distances) fore and aft the plane of sharp focus within which subjects appear sharp in the image.* Subjects outside the DOF appear blurry, having lost all or most fine spatial details (see illustration in [Sec. 1.1](#)). Therefore, to extend the capture volume of iris recognition, we must extend the DOF of the acquisition optics. Indeed, there is demand for large DOF iris acquisition systems [8–10]. However, as shown in Chapter 2, existing solutions using multiple cameras or wavefront coding are often expensive or computationally complex [10] and plagued by noise at higher spatial frequencies [5]. Therefore, we have an immediate need for a solution that significantly extends the DOF of iris acquisition systems, yet have low computational complexity, high signal-to-noise ratio, and be scalable.

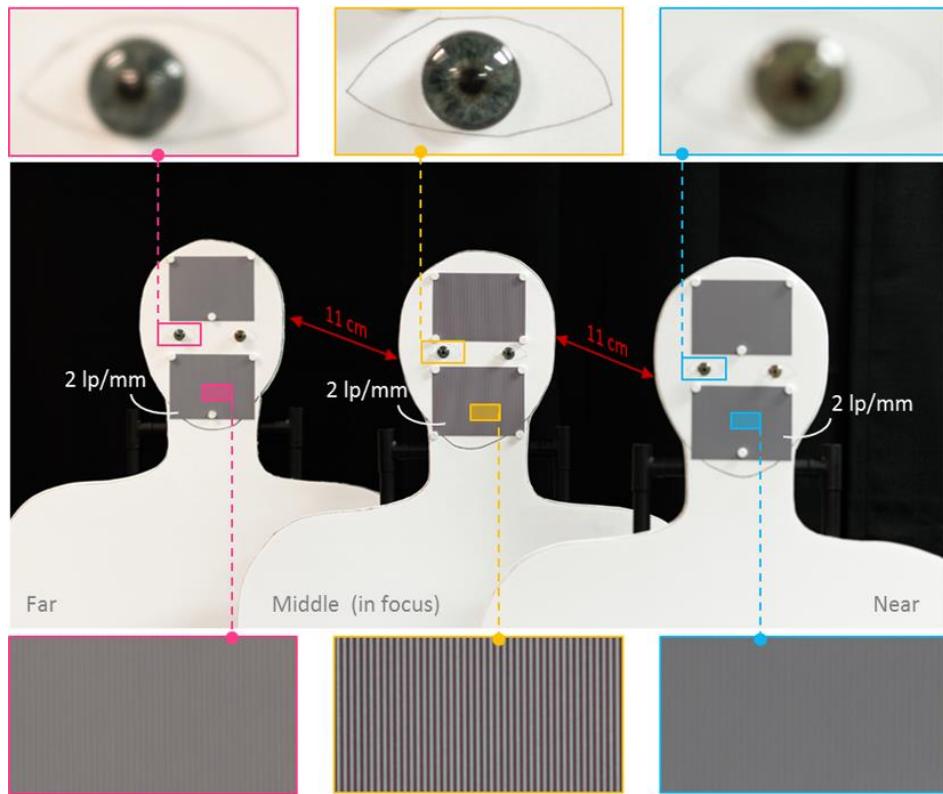
### Problem Statement

The goal of this thesis is to study the DOF problem pertinent to iris acquisition and propose a solution that significantly extends the capture volume. The solution is expected to be scalable, low-cost, and have real-time performance. We restrict the scope to iris acquisition systems only, to optimize over and take advantage of the constraints and the freedoms unique to iris acquisition.

### 1.1 The depth of field problem illustrated

Imposed by the wave nature of light, the DOF limitation is fundamental to all imaging systems.

[Figure 1.1](#) is a photograph of three human-figure cut-outs placed at progressively increasing depths from the camera. We focused the camera on the middle figure cut-out. Observe that spatial details only from the focused cut-out are discriminable in the image. The near and far cut-outs lie outside the DOF. So, the fine spatial details from those cut-outs are lost (blurred) in the image. Recognition degrades when iris images are acquired outside the DOF due to the poor quality of the images.



**Figure 1.1** The depth of field (DOF) problem. Photograph of three human-figure cut-outs with sinusoidal patterns ( $2 \text{ lp/mm}$  in object space) and artificial irises placed apart by  $11 \text{ cm}$  from each other. An  $80 \text{ mm}$  focal length lens with F/5 aperture was focused on the middle cut-out ( $3.6 \text{ m}$  from the lens' vertex). The DOF range is about  $1.3 \text{ cm}$  fore and aft the middle plane (computed using Eq. (1.3) assuming  $c = 6 \mu\text{m}$ ). The spatial resolution in the image falls off with increasing distance fore and aft the middle cut-out.

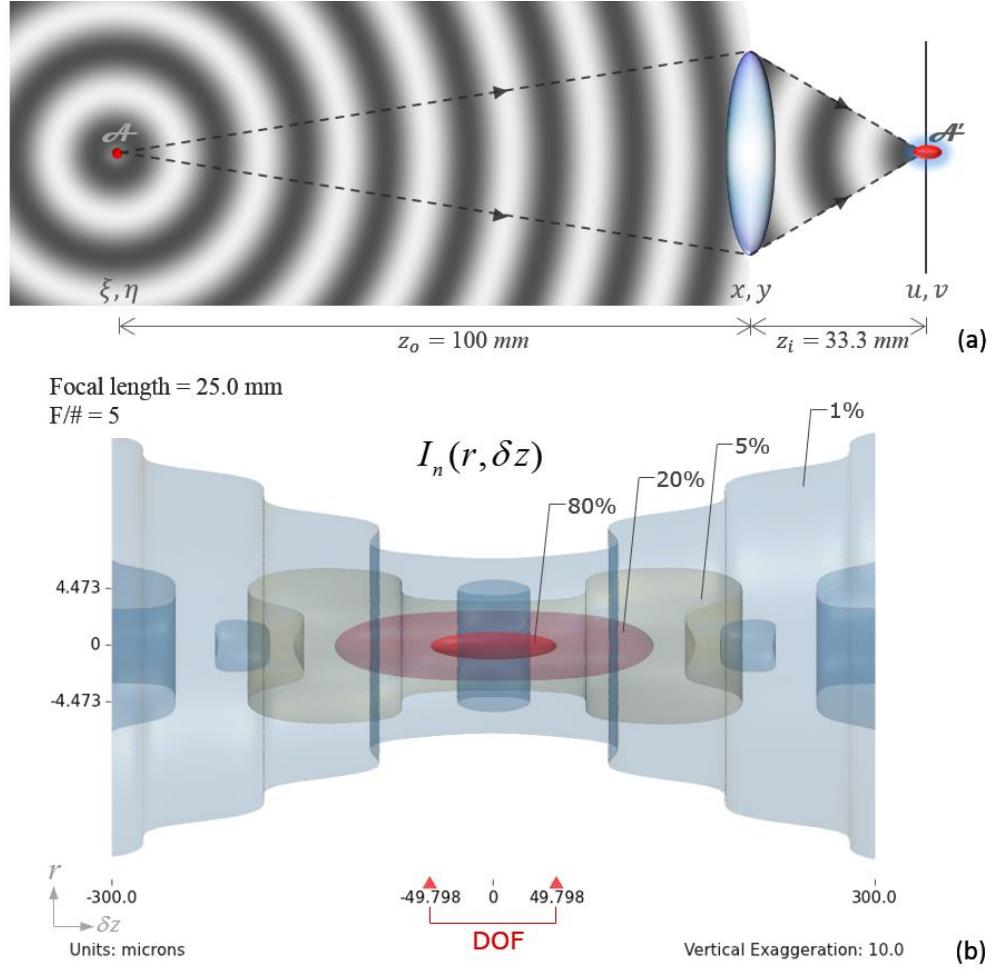
In this work, we have developed a new method called *angular focus stacking* by combining Scheimpflug imaging, a classical photographic technique, and focus stacking, a computational imaging technique. As we will discuss herein, our method is aptly suited for significantly extending the DOF for iris acquisition, simple enough to be easily scalable, and algorithmically fast.

A background on *optical resolution* and depth of field are presented in [Sec. 1.2](#), followed by a background on iris recognition in [Sec. 1.3](#). [Sec. 1.4](#) describes some of the desirable properties of iris recognition systems from the standpoint of imaging. A short background on Scheimpflug imaging and computational imaging are presented in [Sec. 1.5](#) and [Sec. 1.6](#) respectively.

## 1.2 Understanding optical resolution and depth of field

Perfect imaging corresponds to the ability of an imager to produce a scaled replica of an object in the image space [11]. We can think about objects as a collection of independently radiating point sources [11]. A perfect replica can be reconstructed if the complete set of wavefronts emerging from all the point sources is available. However, the finite aperture of the lens can only collect a fraction of the total wavefront. This significant loss of the wavefront in the image space manifests in the loss of high spatial frequency. Therefore, the replica in the image space is never exact.

As shown in [Figure 1.2 \(a\)](#), the image of an infinitesimal point spreads out in space. Using diffraction analysis, we can predict the spread in three-dimension, called the *Point Spread Function* (PSF), if we know the aperture size and the location of the point. The 2D PSF is the cross-section of the 3D PSF in the transverse image plane. For an extended object made of several points, the 2D PSF smears the responses from neighboring points into each other rendering them indistinguishable. The spread along the longitudinal direction limits the ability to discriminate points staggered closely in the direction of the optical axis causing a region of uncertainty. However, the extension of the 3D PSF along the optical axis is the reason why multiple longitudinally separated objects (or points) within a volume in the object space can simultaneously form sharp images. Conversely, an object in the object space may be placed anywhere within this



**Figure 1.2** Incoherent impulse response and DOF. (a) The image  $\mathcal{A}'$  of a point source  $\mathcal{A}$  spreads out in space forming a zone of tolerance called depth of focus (DOF) in the image space; (b) The normalized intensity distribution of the 3D PSF of a 25 mm, F/5 lens imaging an axial point source at a distance of 100 mm. The expression for the 3D PSF was obtained for a circular aperture using scalar diffraction theory and paraxial assumptions. The DOF, having prolate spheroidal shape, is defined as the region within which the intensity has above 80% of the intensity at the geometric focus point. The figure shows iso-surfaces representing 0.8, 0.2, 0.05 and 0.01 intensity levels. The ticks on the left vertical side indicate the locations of the first zeroes of the Airy pattern in the focal plane. The vertical axis has been exaggerated by 10 times to improve the display of the distribution.

---

zone and still form a sharp image. This zone of tolerance in the object space is called depth of field. The corresponding zone in the image space is called depth of focus [12]. We have used the

acronym “DOF” for both the depth of field and the depth of focus. However, the meaning should be apparent from the context. In diffraction theory, the depth of focus is defined as the region of the 3D PSF where the intensity is above 80% of the central maximum [13,14]. This zone is in the shape of a prolate spheroid. In the absence of aberrations, the maximum intensity occurs at the geometric focal point<sup>1</sup>,  $z_g$ , where contributions from all parts of the aperture combine in phase.

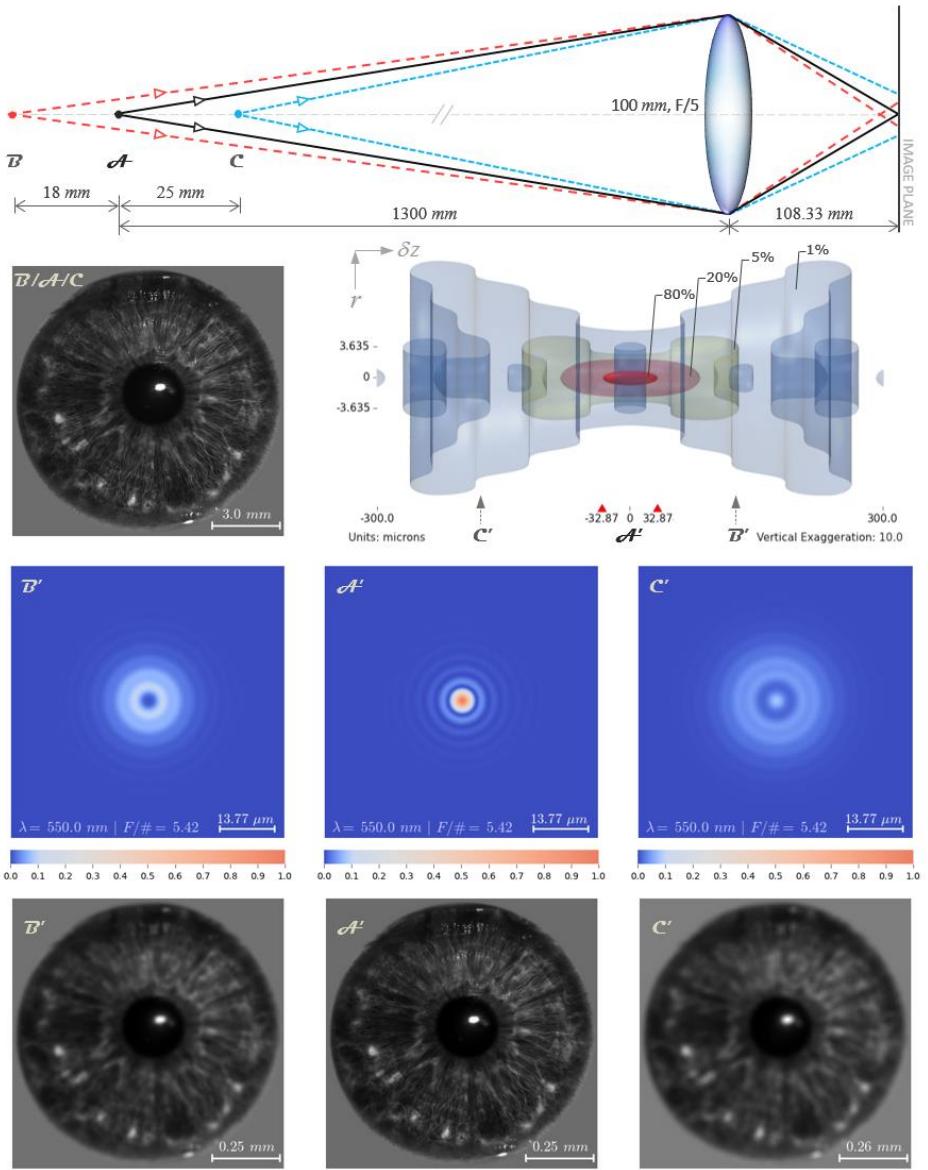
[Figure 1.2 \(b\)](#) shows the aberration-free intensity distribution,  $I_n(r, \delta z)$ , as a function of defocus  $\delta z (= z_i - z_g)$  about the geometric focal point for a point source placed 100 mm from a lens of focal length equal to 25 mm and an aperture diameter of 5 mm. The expression for the distribution—normalized to make  $I_n(0,0)$  equal to unity—was obtained using scalar diffraction theory and paraxial assumptions. The derivation of the expression is shown in [Appendix C.1](#).

The length and breadth of the 80% intensity region, shown in [Figure 1.2\(b\)](#), defines the depth of focus and the transverse spatial resolution respectively and dictates the quality of the image.

A first order optical simulation demonstrating the effect of the DOF in image acquisition at varying depths is shown in [Figure 1.3](#). For this simulation, a 100 mm focal length, F/5 lens focused at 1300 mm is used. In this setup, the imager has a DOF of 9.5 mm in the object space (calculated by applying the lens equation to the extremes of the depth of focus). As it can be seen, the irises located outside the DOF region are severely blurred. Kalka et al. and Sazonova et al. [16,17] showed that the performance of iris recognition deteriorates quickly with increasing amounts of defocus in the captured iris images. Thus, the performance of iris recognition systems operating in a large volume is dependent on the extent of DOF.

---

<sup>1</sup> While this statement is true for imaging systems with Fresnel number ( $a^2/\lambda f$ ) much greater than unity, systems with smaller Fresnel number exhibit a focal shift in the location of the maximum intensity from the geometrical focus point towards the aperture [15].



**Figure 1.3** First order simulation of iris acquisition at multiple depths. The letters **A**, **B** & **C** (**A**, **B** & **C**) denote sources (images). The 1<sup>st</sup> row depicts three sources at three depths from a  $f = 100\text{ mm}$ ,  $F/5$  lens. Point **A** (in focus) forms image **A**. The 2<sup>nd</sup> row shows a  $12\text{ mm}$  iris in the object space (left) and the  $|3\text{D PSF}|^2$  of the source **A** (right). The  $|3\text{D PSF}|^2$  for **B** and **C** are like that of **A** as their relative separation is trivial compared to their distances from the lens. The sensor plane senses the 3D PSF at **A**, **B** & **C** respectively for points **A**, **B** & **C**. The image points **B** & **C**, outside the DOF region in the image space, are  $123.2\text{ }\mu\text{m}$  and  $177.3\text{ }\mu\text{m}$  from **A** respectively. The corresponding incoherent 2D PSFs are shown in the 3<sup>rd</sup> row. The images in the 3<sup>rd</sup> row were obtained by convolving the incoherent 2D PSFs with the de-magnified iris image in 2<sup>nd</sup> row.

Since the early days of photography, landscape photographers such as Ansel Adams have masterfully used Scheimpflug cameras (also called the view cameras) to photograph magnificent vistas with incredible details stretching from the foreground to the receding background (e.g., Ansel Adams, Monument Valley, Arizona, 1958). To capture the intricate details, the masters regularly employed a large aperture (to alleviate the diffraction effects at smaller apertures) and adjusted the orientation of the plane of sharp focus to pass through the subjects in the foreground and the background.

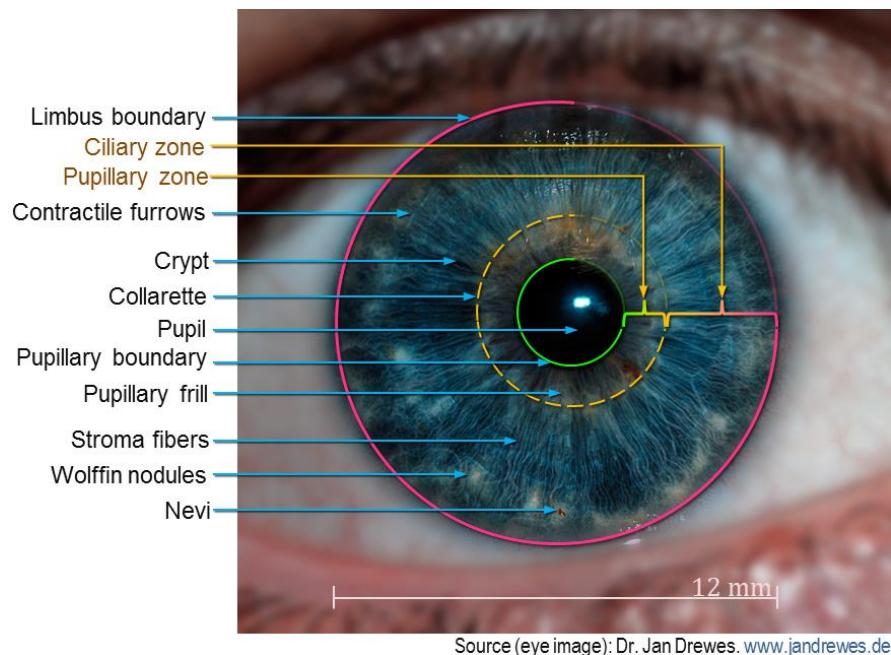
The first step in our approach in solving the DOF problem for iris acquisition is to orient the plane of sharp focus nearly parallel to the ground, such that it passes through the average height of the human eye above the ground. Thus, we reorient the DOF that surrounds the plane of sharp focus to bring the subject's irises across the capture volume within the DOF.

However, the extent of the reoriented DOF may not be sufficient to encompass the required capture volume (along the height). Recently, computational imaging methods such as wavefront coding have been used to transform the light distribution near the focal region for both extending the DOF and improving axial resolution (curtailing the DOF). These class of techniques improves the *instantaneous* DOF of the imaging system. Thus, an extended DOF image can be captured in a single exposure. However, they are usually plagued by noise at higher spatial frequencies. Other computational imaging approaches, such as focus stacking, synthesize an extended DOF image by fusing multiple images captured while varying some parameters of the camera (usually focus).

We posit that an effective way to acquire irises in an extended capture volume is by augmenting modern computational imaging techniques, as a second step, with the classical Scheimpflug imaging technique. But before we delve deep into the aspects of optics and imaging, perhaps it is useful to get a glimpse of what iris recognition constitutes so that we may be able to tailor our solution to fit the needs of iris recognition. A condensed overview of iris recognition follows in the next section.

### 1.3 Primer on iris recognition

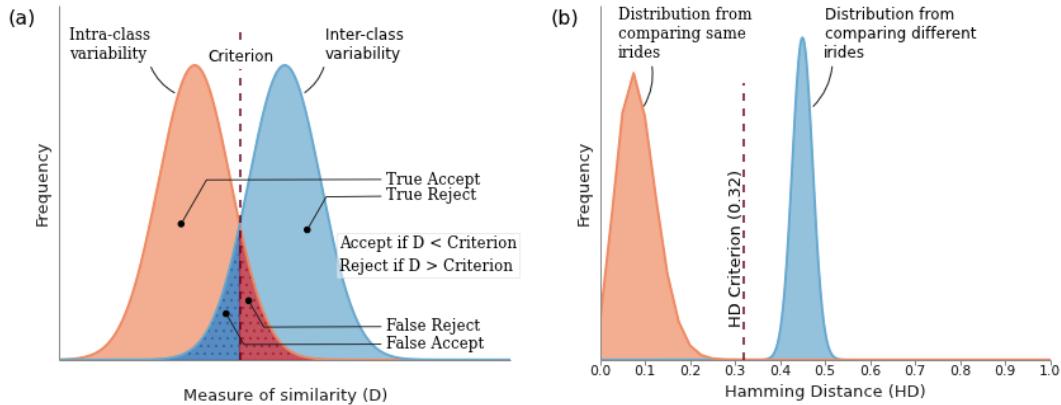
The human iris is the colored portion of the eye having a diameter of 10 mm - 13 mm [2,18]. The iris is perhaps the most complex tissue structure in the human body that is visible externally. The iris pattern has most of the desirable properties of an ideal biomarker, such as uniqueness, stability over time, and relatively easy accessibility. Being an internal organ, the iris is protected from damage due to injuries or intentional tampering [19]. The presence or absence of specific features in the iris is largely determined by heredity (based on genetics); however, the spatial distribution of the cells that form an iris pattern during embryonic development is highly chaotic. This pseudo-random morphogenesis, determined by epigenetic factors, results in unique patterns of the irises in all individuals including that of identical twins [2,20,21]. The diverse microstructures in the iris that manifest at multiple spatial scales [22] are shown in [Figure 1.4](#). These textures, unique to each



**Figure 1.4** Complexity and uniqueness of human iris. Fine textures on the iris form unique biometric patterns which are encoded by iris recognition systems. (Original image processed to emphasize features).

eye, provide distinctive biometric traits. An iris recognition system encodes them into unique templates for identity authentication. Iris color, determined by genetics, is not sufficiently discriminative. Therefore, iris color is not used as a biomarker.

The problem of iris recognition is analogous to binary classification. It involves grouping the members of a set of objects into two classes based on a suitable measure of similarity ([Figure 1.5 \(a\)](#)). Similar objects cluster together as they exhibit less variability within the same class. The figure shows the intra-class variability (randomness within class) and the inter-class variability (randomness between classes) as a function of an appropriate similarity measure,  $D$ . The degree of variability is proportional to the uncertainty of  $D$ . For example, in face recognition, the intra-class variability is the uncertainty of identifying a person's face photographed under varying lighting conditions, poses, time of acquisition, etc. On the other hand, the inter-class variability is the variability between the faces of different subjects. Naturally, the degree of similarity is higher (corresponds to lower uncertainty) in the intra-class grouping.



**Figure 1.5** The iris recognition as a binary classification problem. (a) A schematic of the relationship between the intra-class variability, the inter-class variability and the four possible outcomes. (b) Distribution of the variabilities as a function of Hamming distance for iris recognition. The figures have been adapted from [21].

There are four possible outcomes for the binary classification problem based on the two possible choices in the decision. The region of overlap produces the error rates. The False Accept (or False Positive) is the area of overlap to the left of the decision criterion, and the False Reject (False Negative) is the area of overlap to the right of the decision criterion. If the intra-class variability is less than the inter-class variability [2], then the two distributions are sufficiently separated in space, and the degree of overlap is minimum. We can classify objects reliably in such a case.

[Figure 1.5 \(b\)](#) (from Daugman's paper [21]) is a plot of the distributions of Hamming distances (defined shortly) for 1208 pairwise comparisons of irises from the same eye, and 2064 pairwise comparisons of irises from different eyes. The figure shows that the two distributions are well separated. Thus, we can conclude that Hamming distance measure is robust for the iris recognition problems.

The generation of the iris code broadly consists of four basic steps. An additional matching step is required for the task of identification of a subject based on the iris code [3,23]. A schematic of this pipeline is shown in [Figure 1.7](#). The steps are briefly described below:

1. Iris acquisition

Iris encoding or recognition starts with the *acquisition* of a high-quality image of a subject's eye. Almost all iris acquisition systems use Near Infra-Red (NIR) illumination in the 720-900 nm wavelengths. NIR provides greater visibility of the intricate structure of the iris, which is largely unaffected by pigmentation variations in the iris [2,3,24]. There has also been some evidence of the benefits of using visible illumination for iris acquisition, especially for long-stanoff and unconstrained environments [4].

2. Segmentation and localization

The module following the capture of an acceptable quality iris image is largely known as the *segmentation and localization*. The goal of this step is to determine the spatial extent

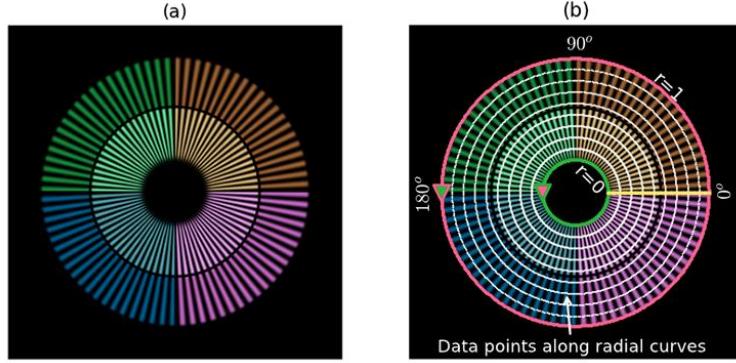
of the iris accurately, locate the pupillary and limbic boundaries and identify and mask out regions within the iris affected by noise such as specular reflections, superimposed eyelashes, and other occlusions that may influence the quality of the template [3]. A wide gamut of algorithms have been proposed for the segmentation and localization of iris regions, such as Daugman's integro-differential operator [2,21], circular Hough transforms along with Canny edge detection [19,22], binary thresholding and morphological transforms [25,26], bit-planes extraction coupled with binary thresholding [27,28], and active contours [29–31].

Once a high-quality iris image has been acquired, segmentation and localization is perhaps the most important steps in biometric template generation. The performance and accuracy of the subsequent stages are critically dependent on the precision and accuracy of the segmentation stage [29].

### 3. Unwrapping/Normalization

The spatial extent of the iris varies considerably amidst captured images due to varying magnification, eye pose, and pupil dilation/expansion [3,21]. Furthermore, the inner and outer boundaries of the iris are not concentric and deviate considerably from perfect circles. Before the generation of the iris code, the segmented iris is geometrically transformed into a scale and translation invariant space. In the new coordinates, the radial distance between the iris boundaries along different directions are normalized between the values of 0 (pupil boundary) and 1 (limbic boundary) [2]. This unwrapping process, shown schematically in [Figure 1.6](#), consists of two steps: (1) several data points are selected along multiple radial curves interspaced between the two iris boundaries, (2) these points, in the Cartesian coordinates, are transformed into the doubly dimensionless polar coordinate system of two variables  $r \in [0,1]$  and  $\theta \in [0,2\pi]$ . Consequently, an iris captured under varying

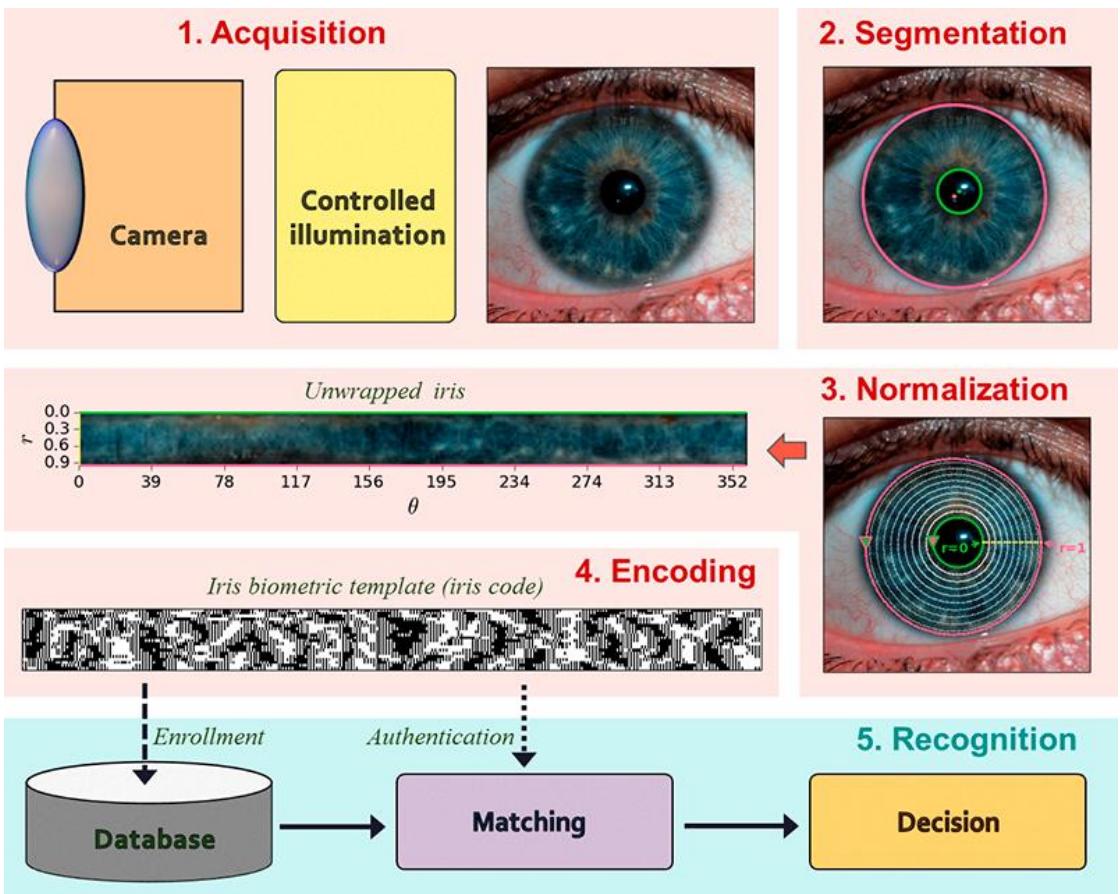
conditions, when compared in the normalized space, will exhibit characteristic features at the same spatial locations.



**Figure 1.6** Schematic of the normalization process using a spoke pattern. (a) A spoke pattern; (b) The green and the magenta circles represents the pupil and iris boundaries, the white markers along the radial curves represent selected points, the yellow line represents  $\theta = 0^\circ$ ; (c) The unwrapped spoke pattern after the normalization process in  $r - \theta$  coordinates.

#### 4. Encoding – generation of the iris code

The encoding process produces a binary feature vector from an ordered sequence of features extracted from the normalized iris image. Following multi-resolution filtering, many iris recognition systems encode the local phase information by quantizing the phase at each location using two bits. Commonly used multi-resolution filters include 2-D Gabor filter [2,10,21], log-Gabor filters [19], multi-level Laplacian pyramids [22], PCA and ICA [30], etc. Like the segmentation module, there are numerous algorithms for iris pattern encoding.



**Figure 1.7** Overview of iris biometric code generation. The different subsystems of iris recognition include the acquisition module, the segmentation module, the normalization module, the encoding module and the recognition module.

##### 5. Recognition – verification/authentication of identity

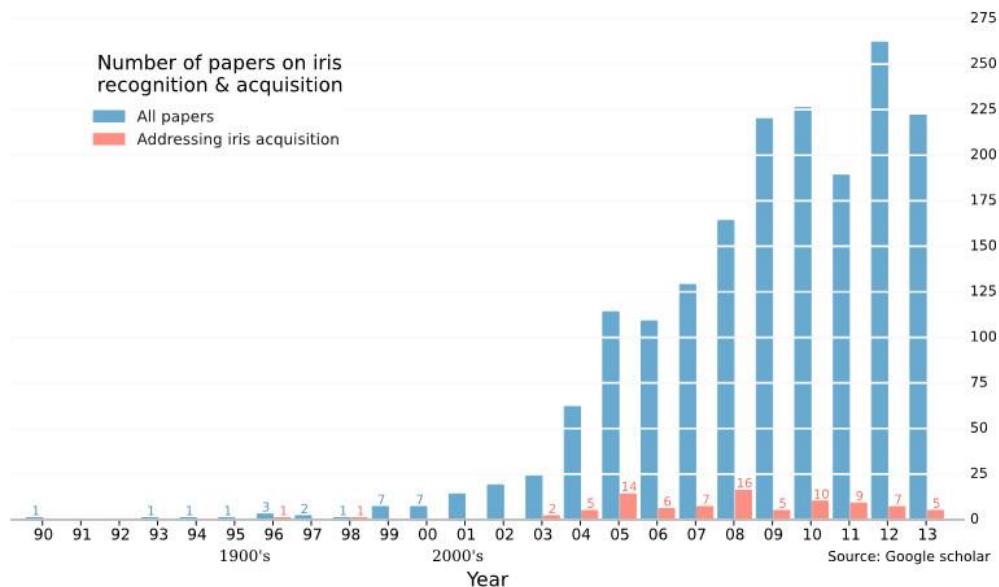
The recognition of iris for subject verification/authentication requires an additional matching step that involves measuring the similarity of a template generated from a newly acquired iris image to one or many templates stored in a database. The most common technique for comparing two iris codes is the normalized Hamming distance (HD), a measure of the percentage of locations where two binary vectors vary [30]. For example, the HD between two orthogonal vectors is 1 whereas the HD between two identical vectors

is 0. However, in practice, due to noise, an HD of 0 between two iris images from the same eye is unlikely. Therefore, an HD of 0.33 or less is used to indicate a match [2].

A detailed exposition of the image processing and pattern recognition techniques employed in the field of iris recognition is beyond the scope of this work. Interested readers are referred to the work of Daugman's [2,29] and the extensive survey of iris recognition by Bowyer et al. [4].

#### 1.4 Desirable properties of iris recognition systems

Iris recognition algorithms have become quite mature and robust in past decade due to the rapid expansion of research both in industry and academia [3,4,29]. A histogram of the scientific publications on iris recognition compiled using Google scholar is shown in [Figure 1.8](#). Despite the



**Figure 1.8** Number of publications in (English) journals on iris recognition between 1990 and 2013. The data was collected using Google scholar by searching the keywords IRIS + RECOGNITION + ACQUISITION + SEGMENTATION + NORMALIZATION + MATCHING. The plot shows that although the total number of research papers on iris recognition has grown tremendously during the last decade, the problems associated with iris acquisition have been overlooked.

unparalleled benefits of iris biometrics and the availability of accurate, robust and state-of-the-art algorithms, iris recognition has not found widespread adaptation as a surveillance technology. One of the main reasons for its low market penetration is the intrusive nature (needing subject cooperation) of most of the currently available systems. These systems, characterized by highly controlled environments, small capture volume, and close range operation, require a significant degree of cooperation from the user. Further, they operate at very close distances to the user, typically within a meter, and often require the user to remain still during the image acquisition. The ease of use, the simplicity of deployment and cost are also critical factors that contribute to the usefulness of iris recognition along with its performance [32]. To make iris recognition less intrusive and easy to use, we need to improve the image acquisition module vastly. The need for better iris acquisition module has been recognized [3,8,33], yet there hasn't been a proportional amount of research in this area as it can be seen in the histogram plot in [Figure 1.8](#).

The accuracy of an iris recognition system is invariably dependent on the quality of iris images captured by the acquisition subsystem. The key design constraints of the acquisition system are spatial resolution, standoff distance, capture volume, residence time, subject motion, subject gaze direction and ambient environment [32]. Three of these constraints are of particular significance to this work, namely spatial resolution, standoff distance, and capture volume. They are described in details in the following paragraphs.

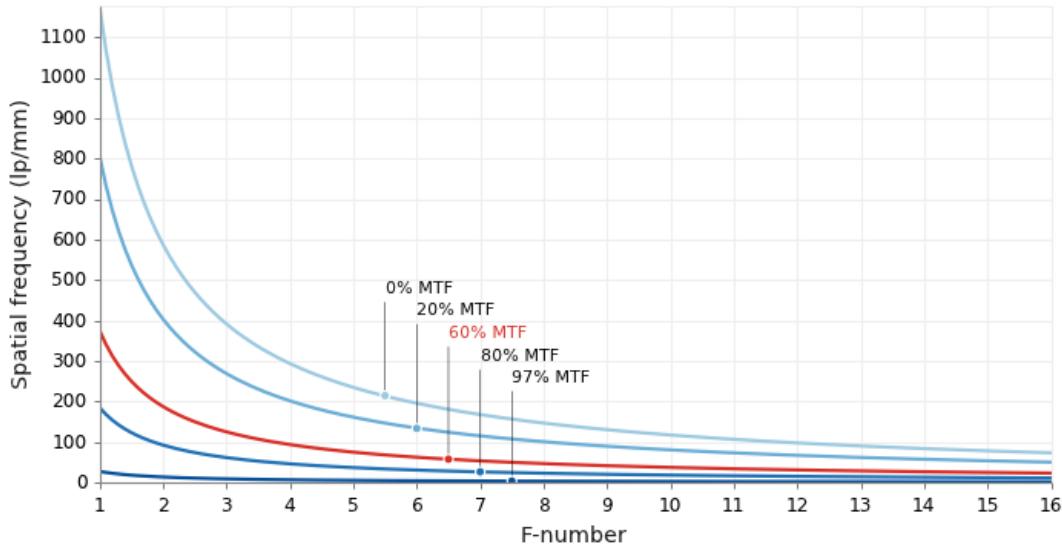
Two types of spatial resolutions are associated with digital imaging systems: the optical resolution,  $r_o$ , of the lens, and the pixel (or sensor) resolution,  $r_p$ , of the digital sensor.

The optical resolution is defined as the maximum spatial frequency that can be resolved in an image at a predetermined contrast [34]. In other words, it is a measure of the ability of an imager to resolve fine details on the object being imaged. The optical resolution is governed by diffraction, and the deviation of the lens from ideal behavior, called aberrations. The resolution at the image

plane of an aberration-free (also known as *diffraction-limited*) lens with *entrance pupil diameter* (EPD),  $D$ , and focal-length,  $f$ , is given by [11]:

$$r_o \approx \frac{D}{\lambda f} = \frac{1}{\lambda F/\#} \text{ lp/mm ,} \quad (1.1)$$

where  $F/\# = f/D$  is the F-number and  $\lambda$  is the illumination wavelength. The optical resolution is measured in cycles per unit-length, typically cycles/mm or line-pairs-per-mm ( $lp/mm$ ). The ISO/IEC 19794-6 [35] standards proposal for MTF recommends iris acquisition devices to maintain a minimum resolution of  $2 \text{ lp/mm}$  at the subject with at 60% contrast [36]. [Figure 1.9](#) plots resolution (maximum spatial frequency) against  $F/\#$  for several contrast values calculated at the image plane for an illumination wavelength of  $850 \text{ nm}$ . The resolution in the object space is obtained by multiplying the image plane resolution by the system magnification. For example, the magnification of a  $100 \text{ mm}$  focal length,  $F/4$  lens imaging at a distance of  $5 \text{ m}$  is about 0.02.



**Figure 1.9** Maximum optical spatial frequency vs. F-number ( $F/\#$ ) for different modulation transfer functions for a wavelength of  $850 \text{ nm}$  at the image plane. The 0% MTF curve corresponds to the diffraction limited cutoff frequency for different  $F/\#$ .

Assuming there are no aberrations, the resolution in the image plane at 60% MTF is 100  $lp/mm$  (the value of red curve at  $F/\# = 4$  in [Figure 1.9](#)). The corresponding resolution in the object plane is  $100 \text{ } lp/mm \times 0.02 = 2 \text{ } lp/mm$ .

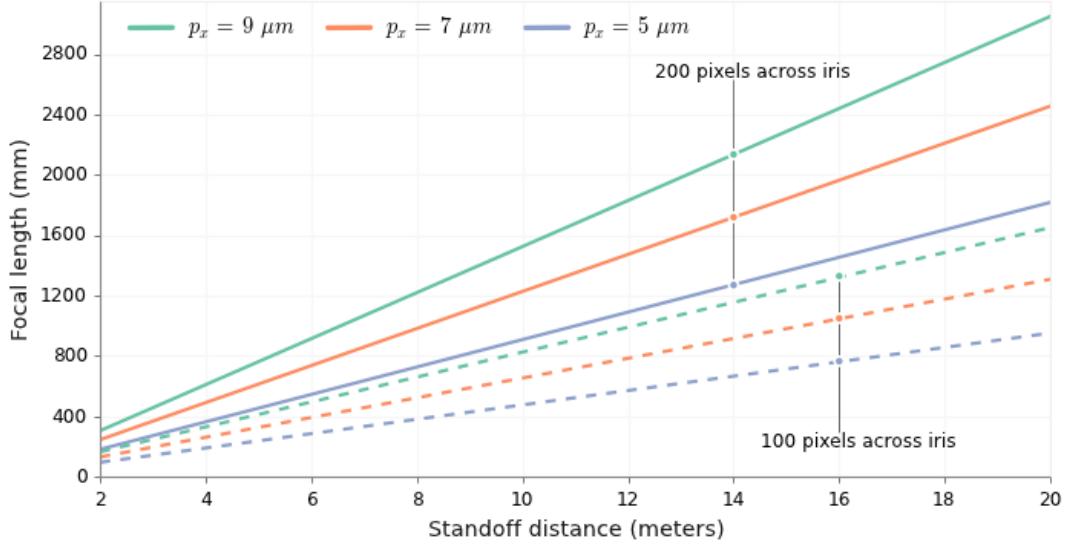
The sensor resolution is determined by the pixel density of the sensor. The ISO/IEC 19794-6 standard [35] requires at least 100 pixels across the iris diameter. Additionally, it recommends a pixel resolution of 200 pixels across the iris diameter [10,32]. For a digital sensor with pixel width  $p \text{ mm}$ , the sensor Nyquist frequency is [37]:

$$r_p = \frac{1}{2p} \text{ } lp/mm . \quad (1.2)$$

The specified number of pixels across the iris diameter also determines the minimum required lateral magnification of the acquisition module [36].

The *standoff* is defined as the distance between the front of the lens and the subject [32]. As suggested earlier, large standoff iris recognition systems are highly desirable. However, capturing high-quality iris images at long distances is a challenging task [32,38]. Increasing the standoff distance while maintaining the pixel count (pixel resolution) on the iris requires the use of higher magnification (higher focal length) optics as shown in [Figure 1.10](#). However, arbitrarily increasing the focal length to form an iris image of 200 pixels does not guarantee adequate optical resolution—an issue that has seldom been discussed in iris recognition literature. Once sufficient sampling has been achieved—either by using high pixel density sensor or through high magnification—the optical resolution ultimately dictates the image quality and consequently has a direct impact on the performance of iris recognition algorithms. As was shown by Ernst Abbe in his treatise on optical imaging, the diffraction limited optical resolution is independent of magnification and is solely determined by the  $F/\#$  [39]. Increasing focal length of the system leads to loss of optical resolution as indicated by Eq. (1.1), unless the lens diameter is increased proportionally to maintain constant  $F/\#$ . However, smaller  $F/\#$  lenses with larger focal lengths tend to be bulky and expensive due

to the use of greater number of optical elements required to correct for aberrations that scales with lens size [40]. Clearly, increasing the standoff distance from a few centimeters to a few meters without significant loss of spatial resolution is a challenge [30].

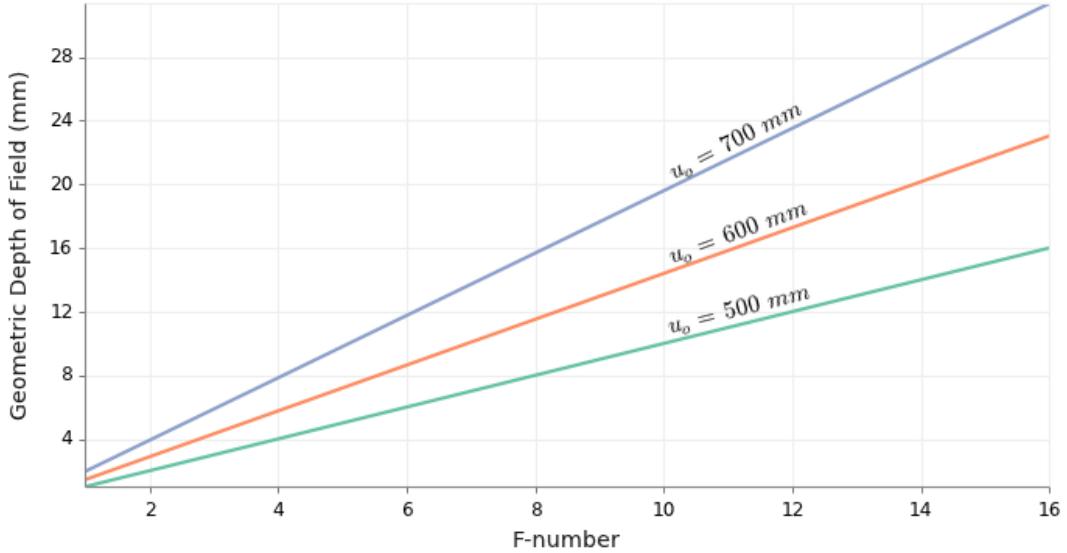


**Figure 1.10** Focal length vs. standoff distance for maintaining 200 pixels (or 100 pixels represented by the dashed lines) across the iris for different pixel pitches. The maximum standoff distance is determined by the focal length (or transverse magnification).

---

For the purpose iris recognition, we define the depth of field (DOF) of the iris acquisition system as the range of object distances within which the spatial resolution required for successful iris recognition is maintained above a predetermined threshold SNR [41]. Incorporating the requirement specified by the ISO/IEC 19794-6 standard [35], this would mean that the DOF is the range of object distances where a spatial resolution of at least  $2 \text{ lp/mm}$  at 60% contrast ratio is maintained. The most frequently employed definition of DOF in iris acquisition literature, derived using geometric optic, is:

$$DOF_{geometric} = \frac{2f^2 F/\# c u^2}{f^4 - F/\#^2 c^2 u^2} , \quad (1.3)$$



**Figure 1.11** Geometric depth of field vs. system F-number ( $F/\#$ ) for various object distances. The geometric DOF linearly increases with the  $F/\#$ . The lens is assumed to have a 50 mm focal length, and the sensor has a pixel width of 5 microns.

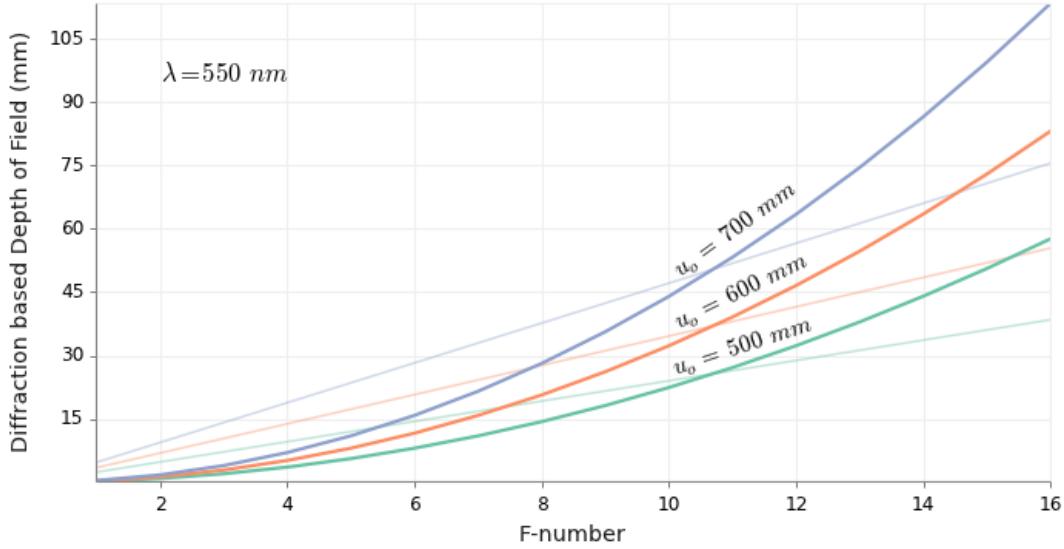
---

where,  $f$  is the focal-length,  $F/\#$  is the F-number,  $u$  is the standoff distance, and  $c$ , the circle of confusion, is a parameter that determines the smallest resolvable feature on the image of an object within the DOF. It is specified by the dimension of the blur spot in the image plane beyond which a point image is ruled out of focus [36]. A plot of the variation of the geometric DOF with respect to the system  $F/\#$  is shown in [Figure 1.11](#). The geometrical DOF defined by Eq. (1.3) increases linearly with the  $F/\#$ .

The notion of the DOF introduced earlier in [Sec. 1.2](#) was defined as the region near the geometrical focus where the intensity drops to 80% of the maximum intensity. That definition is a construct of the scalar diffraction theory [13,14]. The diffraction based DOF is a function of the illumination wavelength and varies with the square of the  $F/\#$ . In the image space, it is given as:

$$DOF_{diffraction} = \frac{12.8\lambda}{\pi} F/\#^2 \approx 4\lambda F/\#^2 . \quad (1.4)$$

[Figure 1.12](#) shows the variation of the diffraction based DOF with respect to the  $F/\#$  for different object distances.



**Figure 1.12** Diffraction depth of field vs. system F-number ( $F/\#$ ) for various object distances. The diffraction based DOF, which are represented by the thick lines, increases proportionally with the square of the  $F/\#$ . The lens is assumed to have a 50 mm focal length, and the illumination wavelength is 550 nm. The thin plots depict the variation of geometric DOF with  $F/\#$  for 9  $\mu\text{m}$  pixel size.

---

The *capture volume* (also called the imaging volume) is typically referred to the three-dimensional spatial volume in which the user's eye must be placed in order to acquire an iris image of predetermined quality [24,32,36]. The lateral extents (measured perpendicular to the optical axis) of the capture volume is determined by the FOV of the camera (provided it has sufficient resolution in the entire FOV), and the axial extent (measured along the optical axis) is determined by the DOF of the lens [36]. Time may also be included as the fourth dimension of the capture volume [10] which specifies the length of time the iris must be placed within the spatial capture volume to capture iris images of sufficient quality and avoid motion blur. For multi-camera systems and systems mounted in pan-tilt units, the net FOV is the total angular extents observable by the acquisition system.

For this thesis, we also define a related term called the *instantaneous capture volume* to distinguish the capture volume of a single/baseline iris acquisition device from the general capture

volume stated by numerous iris acquisition systems employing time and/or space multiplexed cameras.

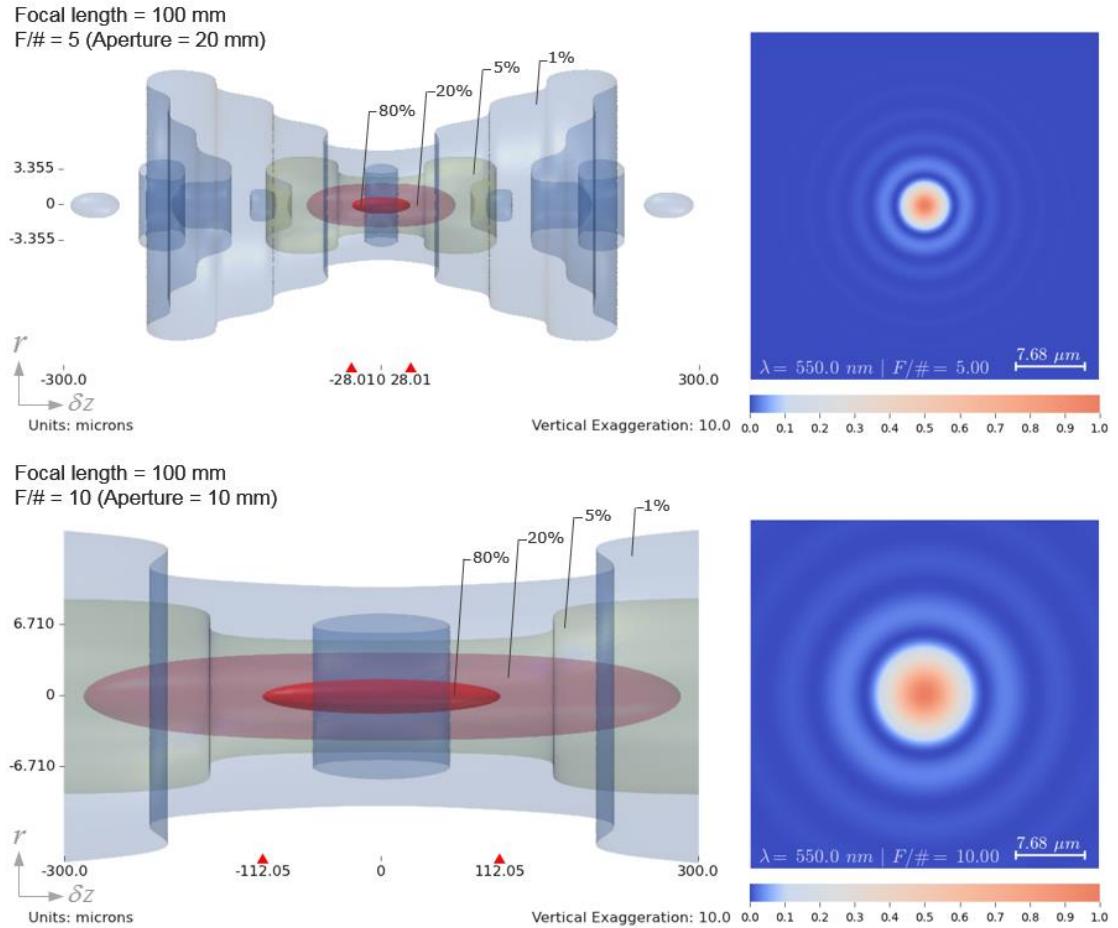
Currently most commercially available iris recognition systems possess shallow DOF resulting in limited usability and increased system complexity [6]. Large capture volumes are not only desirable but also critical for making the iris recognition systems less constraining. Extending the zone of image capture will allow subjects to freely move, albeit while facing the camera, within this zone during the capture process. It will also allow multiple subjects to be identified/verified simultaneously. It has been postulated that increasing the capture volume of current iris acquisition devices significantly will make biometric recognition easier to use and also make them commercially more feasible [7].

The lateral extents of the instantaneous capture volume can be increased by increasing the field of view (FOV). The FOV can be extended by using a larger sensor provided there is sufficient optical resolution at the edges. This implies that the capture volume is fundamentally limited by the DOF. In classical imaging systems, increase in value of the (geometric) DOF is realized by making the system aperture smaller. This relation between DOF and aperture is evident in [Figure 1.11](#) and [Figure 1.12](#). However, stopping down the aperture to increase DOF is not appropriate for iris recognition as it is also accompanied by a loss in optical resolution ([Figure 1.13](#)) and loss of light. Using Eq. (1.4) and the expression for optical resolution Eq. (1.1), the relation between DOF and lateral optical resolution ( $r_o$ ) can be stated as:

$$r_o \approx \frac{4F/\#}{DOF_{diffraction}} . \quad (1.5)$$

Eq. (1.4) also suggests that an  $n$ -fold increase in DOF results in exactly  $n$ -fold loss of light [42]. The loss of light results in a decrease in system SNR.

Various state-of-the-art techniques for increasing the capture volume for iris acquisition along with their advantages and limitations are discussed in Chapter 2.



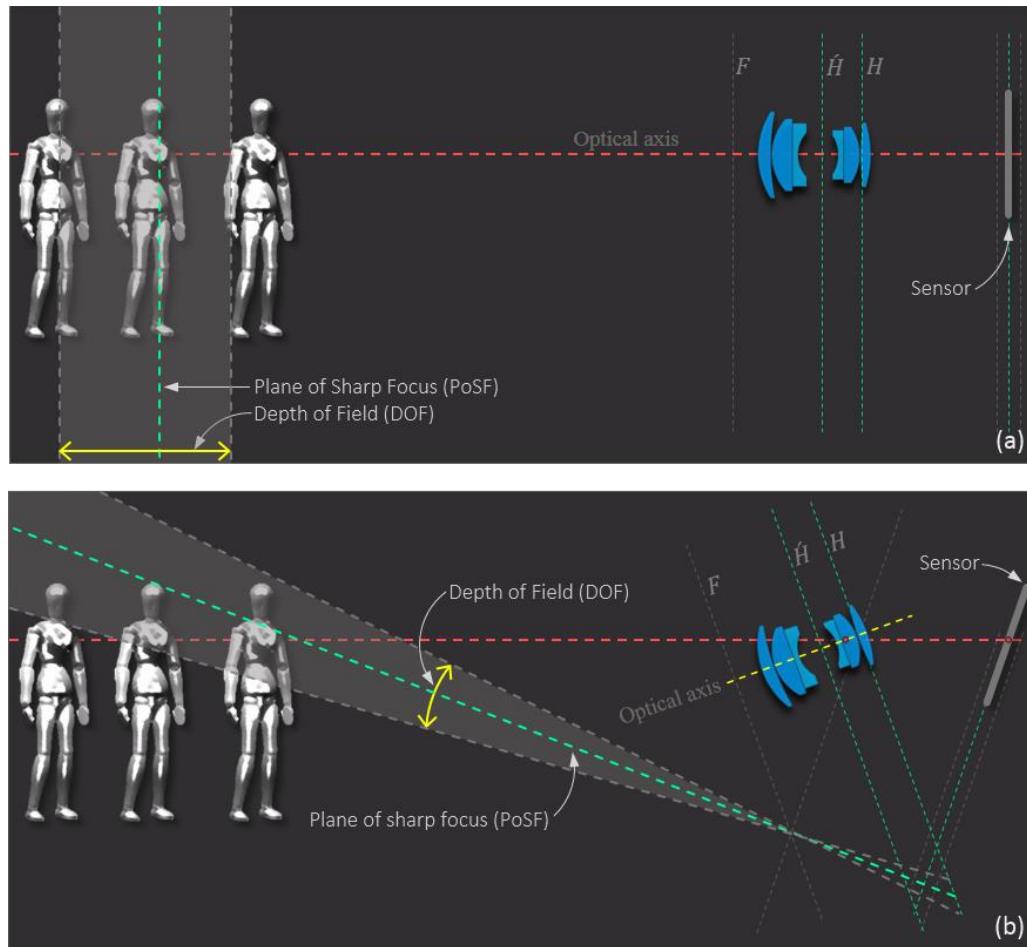
**Figure 1.13** Effect of aperture size on DOF and lateral resolution. The 1<sup>st</sup> row depicts the 0.8, 0.2, 0.05, and 0.01 iso-surfaces of the three-dimensional normalized intensity distribution (left) in the focal region of an aberration-free lens with diameter of 20 mm focused at infinity and the 2D intensity PSF (right) which the cross-section in the focal plane ( $\delta z = 0$ ). The corresponding distributions for a lens with a diameter of 10 mm are shown in the 2<sup>nd</sup> row. We see that halving the aperture diameter (doubling the F/#) increased the DOF four folds as indicated by the red triangle markers on the horizontal axis; however, the diffraction spot-size also increased two folds, consequently halving the spatial resolution.

## 1.5 Scheimpflug imaging

In a conventional camera (rigid) the cardinal planes of the lens are parallel to the sensor plane as shown in [Figure 1.14 \(a\)](#). The cardinal planes are planes perpendicular to the optical axis and passing through the cardinal points—the front and rear focal and principal points. The plane of

sharp focus (PoSF) in the object space in the rigid camera is parallel to the lens' cardinal plane and sensor planes. This prevalent imaging configuration in which the PoSF, the lens planes and the sensor plane are parallel is called the *frontoparallel* configuration. In contrast, as shown in [Figure 1.14 \(b\)](#), the sensor plane and the lens planes in a Scheimpflug camera free to rotate about independent pivots. The rotation of the lens and/or the sensor induces a corresponding rotation of the PoSF [43,44]. Geometric models of Scheimpflug imaging often idealizes the lens as a thin surface with certain power. In such thin-lens models, the front and rear principal planes coalesce in to a single plane passing through the center of the lens. This plane passing through the center of the thin lens is designated as the lens plane. In thick lens models that represent the lens using the two principal planes (along with the cardinal points), there are essentially two lens planes—the front and rear principal planes as shown in [Figure 1.14 \(b\)](#). In fact, we have introduced a new model for Scheimpflug imaging in this thesis (Chapters 3 and 4) wherein the lens planes are the extensions of the paraxial entrance and exit pupils. In the rest of the thesis the phrase lens plane, unless explicitly qualified, shall mean the two planes collectively in a thick lens model or the single plane in a thin lens model.

Since the PoSF, lens, and the sensor planes are no longer restricted to remain mutually parallel, their orientations may be tailored to optimize for a desired imaging goal. For example, architecture photographers regularly employ the sensor tilt and shift to correct for perspective distortions, landscape photographers use the lens and/or sensor tilts to orient the PoSF at some angle that maximizes the focus across a deeply receding subject [44]. Scheimpflug cameras have also been used to extend the depth of field in scientific and medical imaging fields such as for range estimation [45], particle image velocimetry [46], and retinal imaging [47].



**Figure 1.14** Frontoparallel vs Scheimpflug imaging. (a) In frontoparallel configuration the sensor plane is parallel to the cardinal planes (the planes perpendicular to the optical axis and passing through the cardinal points:  $F, H, \tilde{H}$ ). (b) In Scheimpflug imaging, the lens and the sensor planes are free to rotate about independent pivots.

## 1.6 Computational imaging

By orienting the plane of sharp focus (PoSF) in the object space at some advantageous angle (for example, it may span horizontally across the imaging volume at the average height of the subject group), we can gather focused images of the iris within the depth of field (DOF). However, with small  $F/\#$  (to improve SNR and optical spatial resolution), the extents of the DOF many not be sufficient. Therefore, to further extend the limits of the DOF, we may employ computational

techniques. By *computational techniques*, we mean that the final iris image, which is used for iris recognition, is not the digital image at the sensor plane, but rather an image that is synthesized algorithmically to maximize the resolution of the iris image.

Typically, a computational imaging system is a hybrid system consisting of both hardware (for example, the lens elements, digital sensor, amplitude and/or phase aperture masks, shutter systems, etc.) and computational algorithms that are jointly optimized for gathering the maximum possible information encoded in the object field [48]. The data obtained from the digital sensor may not even represent a typical image of the object as obtained in a conventional camera.

Computational algorithms, typically some form of inverse imaging or reconstruction procedure, are an integral part of the imaging pipeline that is used to synthesize the final image. In [Sec. 2.3.2](#), we provide a survey of computational imaging approaches used to overcome the DOF problem in the context of iris recognition systems. In this thesis, we have adopted a multi-frame image reconstruction approach to computationally synthesize an image exhibiting an extremely large DOF from a sequence of images captured while rotating a lens. The main advantage of our simple method is that the computational complexity of the reconstruction is extremely low and ideal for implementing in real-time applications.

## 1.7 Summary

In this chapter, we introduced the problem of depth of field in iris acquisition systems. To provide sufficient context for the problem at hand, we provided a short overview of the iris biometric technology and discussed the desirable properties of iris acquisition systems. We followed up the brief foray into the background knowledge with quick introductions to Scheimpflug imaging and Computational Imaging—the two tools we will weld to expand the imaging volume of iris recognition systems.

## Chapter 2

### STATE-OF-THE-ART

*[...] steal from one [...], it's plagiarism; [...] steal from many, it's research.*

—Wilson Mizner

This chapter provides a survey of current and past efforts that aim to extend the capabilities of iris image acquisition systems. The survey not only includes research that were designed to solve the depth of field (DOF) problem, but also those investigations that have tried to capture iris images at significantly large standoff distances. We have presented a comparison of the features relevant to iris image acquisition in Table 2.1. A visual representation of the capture volumes of selected systems from Table 2.1 is shown in Figure 2.1. It is in light of the trials, tribulations, and triumphs of these past efforts that we can discern the desire and difficulty of resolving the problem at hand.

#### 2.1 State-of-the-art large standoff iris acquisition

One of the earliest systems for iris recognition at a significant distance was reported by Sarnoff [49] in 2005. It used two custom designed elliptical mirror telescopes with diameters of 101.6 *mm* and 203.2 *mm* to demonstrate the feasibility of iris recognition at large standoff distances of 5 *meter* and 10 *meter* respectively. The magnifications in their telescopes were sufficient to produce iris images with a pixel resolution of 128 pixels for stationary subjects. They used a Hitachi KP-M2RN infrared camera that has a pixel pitch of 9  $\mu\text{m}$ . Since the average iris diameter is 11 *mm*, back-of-the-envelope calculations (not published in the paper) suggest a focal length of around 585 *mm* and

1170 *mm* respectively for operating distances of 5 *meter* and 10 *meter*. We can also infer the geometrical DOF, using Eq. (1.3), to be around 7.6 *mm*. To capture quality iris images at such low DOF, the subjects' heads were supported using a chin rest and the image acquisition setup was constructed on a moderately large optical table [24]. Further, to improve the odds of obtaining at least one good quality iris image, their system captures a sequence of video frames at 12 *fps*. The measured system Modulation Transfer Function (MTF) in the object space for both distances indicated a spatial resolution of 2 *lp/mm* at 80% modulation strength. Since both the *F/#* and the magnification were kept constant for both 5 *meter* and 10 *meter* distances, the optical performance in terms of spatial resolution was adversely impacted by doubling of the distance. Although, such a system is hardly suitable for deployment, the study demonstrated the feasibility of iris recognition at large standoff distances.

Sarnoff built another long-distance iris acquisition system in 2007 using a Meade LX200-R F/10, 8 *inch* reflecting telescope operating at 850 *nm* that generated useful images over 15 *meters* [24].

A 18 *meter* standoff distance iris acquisition system that incorporated adaptive optics was demonstrated by AOptix™ in 2007 [24].

The HBOX™ from Global Rainmakers [50] provided the ability to capture multiple subjects walking through a portal system at a distance of about 1.5 *meters*.

The Iris On the Move (IOM) [10,32] portal system developed by Sarnoff in 2008 is capable of acquiring iris images at distances of approximately 2 – 3 *meters* with a subject walking at normal pace. It uses an off-the-shelf 210 *mm* lens and a camera with a pixel pitch of 7.4  $\mu\text{m}$ . The IOM system uses a focusing mechanism along with video capture to acquire iris images. The reported capture volume is at least 100x100x5 *mm*<sup>3</sup> for a pixel resolution of 200 *pixels/cm* and optical resolution good enough for iris recognition. Also, the DOF was not calculated based on a given definition of the circle-of-confusion, but rather found empirically by identifying the region wherein

the recognition algorithm produced a Hamming distance below 0.33 when matching templates and iris codes from the same eye. A general disadvantage of portal based systems is that the user movement is restricted to a narrow zone. Such restrictions may prevent these systems from being deployed in crowded locations.

A combined face and iris recognition system (CFAIRS) developed by Honeywell in 2008, which uses multiple field of view (FOV) cameras, is capable of acquiring face and iris images of multiple subjects within a range of 1 to 5 m [51]. The Retica's Eagle-Eyes<sup>TM</sup>, another multi-biometric system that uses multiple cameras in addition to video tracking, has a capture volume of 3x2x3 m<sup>3</sup> and a standoff distance of 3 m [52,53].

An iris recognition system at a standoff distance of about 3 m was shown by Dong et al. in [54]. It uses a 300 mm lens set at F/15 mounted on a pan-tilt unit and has a capture volume of 20x15x10 cm<sup>3</sup> for acquiring iris images with greater than 150 pixels across.

Boehnen et al. [38,55] demonstrated the Standoff Multimodal biometric system that is capable of acquiring iris images from a distance of 7 m. It uses a multi-camera approach to increase the capture volume; the iris camera is a 150 – 500 mm, F/5 – F/6.3 zoom lens.

More recently, a large standoff iris acquisition system with a pixel resolution of 200 pixels on the iris at a distance of 4 – 8 m using an 800 mm COTS auto-focus camera has been demonstrated by Venugopalan et al. in [56].

Stoker et al. demonstrated an outdoors iris recognition system operating at a distance exceeding 25 m using a 10 inch (254 mm) Ritchey-Chretien telescope, fitted with several custom optical elements for aberration correction and collimation [57]. Additionally, image restoration techniques were used to improve the quality of images degraded by atmospheric turbulence.

## 2.2 State-of-the-art iris acquisition with large capture volume

The total capture volume in almost all state-of-the-art iris acquisition systems can be dynamically improved using a combination of one or more techniques such as: (1) Wide-Field-Of-View

(WFOV) and Narrow-Field-Of-View (NFOV) cameras [51,58–60], (2) Pan-Tilt-Zoom (PTZ) camera system [53,55,58,60–63], (3) multiple stacked cameras [52,53], (4) tracking [56,64], and (5) video/multiple-capture [10,52,53]. A stationary WFOV camera is frequently used in such systems for initial detection and tracking of one or more subjects' faces. Once the position of a subject's face is estimated either by using an auto-focus mechanism, or by stereo techniques [60,61], or through active illumination [63], an NFOV camera mounted on a Pan-Tilt-Unit (PTU) is used to capture a magnified image of the iris. PTZ systems can dynamically adapt the focus, look-direction, and zoom of the camera to accommodate the subject.

A prototype “stand-off” iris recognition system described by Wheeler et al. [58] uses two calibrated WFOV cameras for detecting the subject’s face. A separate camera (having a pixel size of  $6.45 \times 6.45 \mu\text{m}$ ) sensitive to NIR illumination with motorized focus and zoom lens ( $10 - 160 \text{ mm}$  focal length) mounted on a PTU is used for iris capture. With an optimum aperture setting of F/9.5, the reported DOF of the iris camera is about  $30 \text{ mm}$ . They captured images with 200 pixels across the iris diameter for multiple subjects from a standoff distance of  $1.5 \text{ m}$ .

The system developed by MERL [60] uses a WFOV video camera (to detect faces) coupled with a high-resolution COTS digital still-camera (pixel size of about  $7.4 \mu\text{m}$ ) with  $70 - 200 \text{ mm}$  zoom lens mounted on a PTU to capture iris images within a distance of  $1.2 \text{ m}$  to  $2.1 \text{ m}$ . They estimate the subject distance from the camera from the facial features. Two major drawbacks of this system are that face detection based scanning is time-consuming and distance estimation based on *average* facial features is sub-optimal [65].

Yoon et al. [63,65] developed an iris acquisition system consisting of a WFOV camera and a PTZ camera that is guided by a light stripe projector. The capture volume, between  $1.5 \text{ m}$  and  $2.5 \text{ m}$ , was improved using a  $70 - 300 \text{ mm}$  zoom lens attached to the PTZ camera. The system could capture iris images with a pixel resolution of more than 150 pixels across the iris diameter.

Further, they project an NIR light plane onto the scene to aid in the estimation of the exact location of the subject.

### **2.3 State-of-the-art iris acquisition with large *instantaneous* capture volume**

#### **2.3.1 Extending capture volume using image processing**

Image processing based methods for extending the region of image capture consists of either deconvolution techniques that are often formulated in a Bayesian framework [66,67], or image restoration techniques such as the constrained least square (CLS) filtering [68–70]. Deconvolution based methods invariably involve the estimation of blur kernel and defining suitable priors. The blur kernel is frequently estimated from the knowledge of the subject distance from the camera using simple geometrical techniques when the focal length is known. The priors are often computed from iris image statistics and gradient properties of natural images, which exhibit heavy-tailed distributions. The CLS is a robust inverse filtering technique, like the Wiener filter, that requires the knowledge of the degradation function (generally the defocus PSF of the lens) and an appropriate choice of a regularization parameter.

Although the papers mentioned above, claim to have extended the usable region over which iris acquisition is possible, no technique based solely on digital processing of images can improve the spatial resolution or optical bandwidth of the system. Moreover, all the above methods assume either a Gaussian or a pill-box model for the PSF which is considerably different from the optical PSF.

#### **2.3.2 Extending capture volume using Wavefront coded systems**

The transfer function corresponding to a defocused PSF (image of a point source that is not in focus) exhibit nulls—points at which the function becomes zero. The nulls in the transfer function pose problems to the inverse filtering. Extended DOF is achieved in wavefront coded system by rendering the system’s PSF invariant to defocus blur. A phase mask is often placed at the location of the stop or, at the position of any one of the two pupils. The mask modifies the impulse response

to become depth invariant. Consequently, the transfer function is rendered free of nulls. Thus, at least in theory, a single de-blurring filter can be used to recover high fidelity iris images over a large range of distance in the object space [5].

Plemmons et al. [6,9] showed that the DOF of iris recognition systems can be improved by a factor of six (from  $5\text{ cm}$  up to  $30\text{ cm}$ ) by wavefront coding using cubic phase masks whose phase profile is separable in rectangular coordinates. Their system consisted of an F/8 lens with effective focal length of  $57\text{ mm}$  and a digital sensor with  $6.7\mu\text{m}$  pixel size. The best-focus distance of the system was at  $500\text{ mm}$ .

In [34], Narayanswamy et al. demonstrated the benefits of using custom wavefront-encoding optical elements specially designed for iris biometrics in which they improved the DOF by three folds at a standoff distance of  $0.55\text{ m}$ . Subsequently, they used a task-optimized wavefront coded system to increase the standoff distance up to  $2\text{ m}$  while improving the SNR of the system compared to the previous wavefront coded solutions [41]. However, improvement in the DOF of wavefront coded systems often comes at the cost of lower SNR in the higher spatial frequencies. The performance of the wavefront coded systems for information acquisition at higher spatial frequencies can be improved by the use of more general, non-separable phase profiles as demonstrated Barwick in [5]. Furthermore, in [8], Boddeti and Kumar showed the use of linear correlation filters for iris recognition while achieving a four-fold improvement in DOF. Instead of using the Hamming distance as a measure of similarity between the iris templates, their method relies on the degree of correlation between the un-restored intermediate iris image and a template generated from the same eye using the wavefront coded system. The direct use of the optically encoded intermediate images is particularly attractive since it avoids the often computationally complex digital post-processing involved in restoring a high-fidelity image from the intermediate image. Recently, a wavefront coded,  $400\text{ mm}$ , F/6.3 lens combined with PTZ camera was shown to improve the DOF from about  $6\text{ cm}$  to  $20\text{ cm}$  at a standoff distance of  $3\text{ m}$  in Hsieh et al. [71].

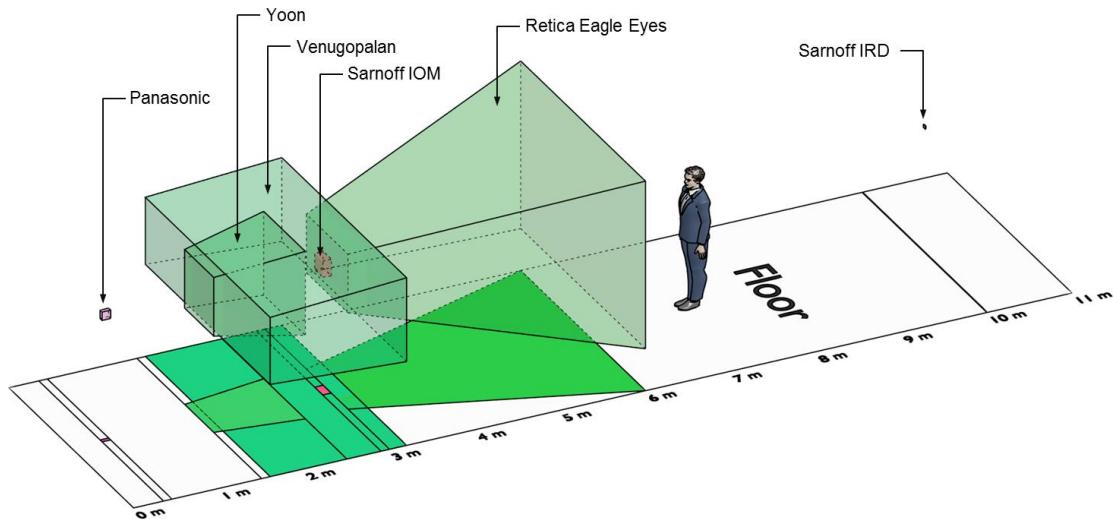
**Table 2.1** Comparison of features in the state-of-the-art iris acquisition systems.

System	Standoff (m)	Capture volume (HxWxD) cm <sup>3</sup>	Pixels on Iris	Focal length (mm)	F/#	Techniques to improve capture volume/standoff
LG-3000	0.1	2x2x10				
Panasonic BM-ET 300	0.35	10x5x10				Video
IrisGuard IG-H100	0.4	5x5x15				
OKI IrisPass-WG	0.45	18x55x30				
Sarnoff IRD	5 / 10	3.69x4.92x0.76*	128	585 / 1170*		Custom built telescope, video capture
Sarnoff	> 15				F/10	Meade telescope
Sarnoff IOM	2 – 3	20x20x10	100	210		Focusing, video capture (multi- modal)
Honeywell's CFAIRS	1	-x-x400				Multiple cameras (multi- modal)
Retica Eagle- Eyes™	3	300x200x300				PTU, multiple cameras (3)
Guo et al. MERL	1.2 – 2.1	-x-x60		70 – 200		PTU, multiple cameras, zoom lens
Wheeler et al.	1.5	-x-x3	200	10 – 160	F/9.5	PTU, multiple cameras
Dong et al.	3	20x15x10	150	300	F/15	PTZ, automatic audio feedback, tracking, customized focusing

Boehnen et al.'s Standoff Multimodal	7			150 – 500	F/5.6 – F/6.3	PTZ, multiple cameras (3), telephoto zoom lens (multi-modal)
Venugopalan et al.	4 – 8		200	800		Long telephoto lens
Venugopalan et al.	0.5 – 1.6	-x-x160	200 – 100	119		PTZ, video
Stoker et al.	25			254		Telescope fitted with custom optics, image restoration
Yoon et al.	1.5	100x100x100	150	70 – 300		PTZ, multiple cameras
Plemmons et al.	0.5	-x-x30		57	F/8	Wavefront coding
Hsieh et al.	3	-x-x20		400	F/6.3	Wavefront coding, telephoto lens, PTZ

\* indicates an estimated value in the cited paper. The actual value may not have been reported by the cited paper.

Wavefront coding provides a viable solution for increasing the instantaneous capture volume for iris recognition without reducing the system aperture albeit, at the cost of lower SNR for higher spatial frequencies and greater computational complexity. Wavefront coded iris recognition systems are also susceptible to the effects of specular reflections, which are amplified during the wavefront coding and throws off the segmentation algorithm.



**Figure 2.1** A visual representation of the imaging volumes of selected systems from Table 2.1. The imaging volume's projection on to the floor shows the it's width and standoff distances. The imaging volumes of systems which use PTZ and multi-camera arrangement are shown in shades of green.

(An orthographic camera viewpoint has been used in the rendition to avoid distortions due to perspective foreshortening. The human model in the figure is licensed under a Creative Commons Attribution 3.0 United States License and is Copyright © 2003-2013 Andrew Kator & Jennifer Legaz.)

---

## 2.4 Summary

We looked at the several methods used for extending the capture volume of iris acquisition systems. Most purely opto-mechanical systems use a combination of wide field of view (WFOV) and a PTZ mounted long focal length, high magnification narrow field of view (NFOV) lens to increase the capture volume. Almost all the hybrid computational imaging techniques employ wavefront coding to make the PSF invariant to defocus blur, extending the range over which usable iris images can be obtained. However, extended DOF in these computational imaging systems comes at the cost of lower signal-to-noise ratio and computational complexity.

## Chapter 3

### MODEL OF SCHEIMPFLUG IMAGING – I: PROPERTIES OF IMAGE

*Essentially, all models are wrong, but some are useful.*

—George Box

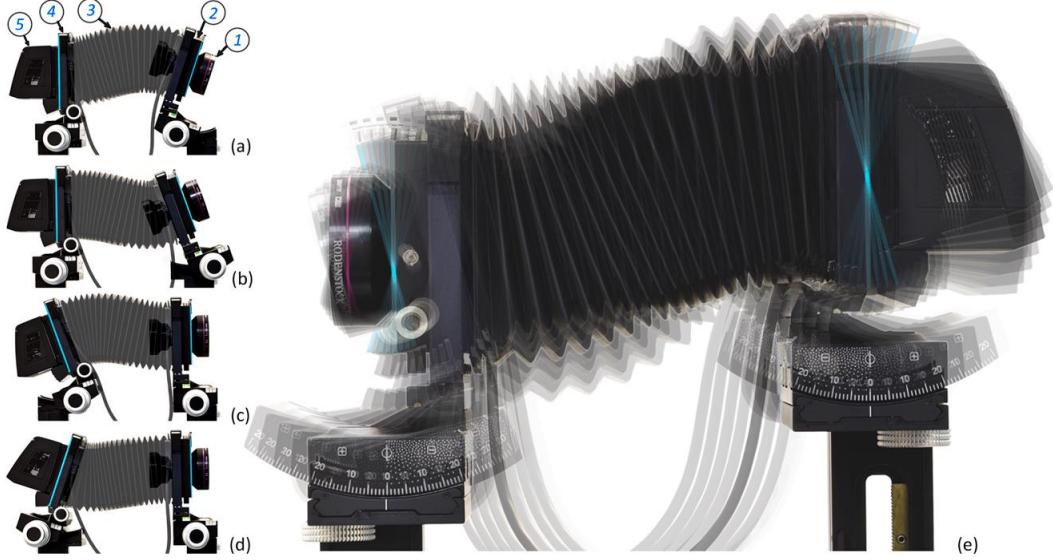
For our investigation of the use of a Scheimpflug camera for extending the imaging volume of iris cameras, we require a model that includes the essential characteristics of Scheimpflug imaging. Although there are existing models, several of them use the thin lens approximation that is oversimplistic. For example, it is impossible to accurately describe the shift in the image field due to lens rotation using a thin lens model. On the other hand, models that do use thick-lens abstraction, while accurate, do not explicitly consider the effects of the pupils (defined later) on image formation. The lack of pupil parameters in these models makes it hard to predict the geometric properties of the image obtained using a Scheimpflug camera in which the lens and image planes are free to rotate about independent pivots. Therefore, we develop a new model that depicts the explicit dependence of the pupils on the properties of the image in Scheimpflug cameras. We have broken down the modeling process into two chapters—in Chapter 3 (this chapter) we develop the relationship between a world point and its image in a Scheimpflug camera, verify the model using optical ray tracing in Zemax, and finally study the consequence of rotating the sensor and lens on

the geometric properties of the image using our model. In Chapter 4 we derive object-image equations that relates the angles and directed distances between the object, lens and sensor planes for which the object plane is brought to focus on the sensor plane.

Conventional imaging systems consists of a lens and an *image plane* on which a sharp image of an *object plane* is formed. The object and image planes are called *conjugates*. Further, amongst the several possible planes that are perpendicular to the optical axis and pass through the lens, we designate the one as the *lens plane* that provides some advantage in the geometric model. For example, the plane through the lens center and the plane through the object-side principal point are designated as the lens planes in the thin-lens and thick-lens models respectively. In our model, the lens plane is the plane through the center of the entrance pupil (defined later).

The plane in the object space that is in sharp focus is called the plane of sharp focus (PoSF). In the *fronto-parallel* configuration used in most camera designs, the lens and the sensor planes are parallel to each other and perpendicular to the optical axis. In such designs, the physics of optical imaging—described by the *Gaussian Formula*—dictates that the PoSF must be parallel to the lens and image planes. In contrast, the lens and image planes in a Scheimpflug camera are free to swivel about their pivots (as shown in [Figure 3.1](#)) resulting in a corresponding swivel of the PoSF. We exploit this feature—the freedom to arbitrarily orient the PoSF—of Scheimpflug imaging to improve the depth of field of the iris acquisition devices. However, the degrees of freedom offered by the Scheimpflug camera comes at the cost of added complexity of operation. Therefore, a rich description of such cameras requires the development of a general model. In this chapter, we initiate the development of such a model of Scheimpflug imaging starting from the axioms of *geometric optics (ray optics)*.

Assumptions are crucial and necessary for modeling that enable its expediency but limits its applicability. To make the problem tractable, we assume paraxial imaging, rotational symmetry and aberration-free optics. Additionally, we assume the refractive index of the lens elements and



**Figure 3.1** Scheimpflug camera movements. Insets (a), (b), (c), and (d) depict key camera movements—lens and sensor plane tilts about a horizontal axis—amongst several possible, in a Sinar P3 camera. Labels indicate the lens (1), the lens standard (2), bellows (3), sensor standard (4), and the sensor (5). The cyan lines on the two standards accentuates the orientations. Inset (e) is a superimposed sequence of images of the camera with the two standards in a multitude of orientations. The physical locations of the two pivots emerge at the intersection of the superimposed cyan lines.

---

the interstitial medium to be isotropic (uniform along all directions) and homogeneous (uniform at all positions); this assumption imposes rectilinear propagation of light. Further, we assume the lens is surrounded by air (of refractive index equal to one). Consequently, the front and back focal lengths are equal, and the two nodal points coincide with the corresponding principal points.

In the following sections of this chapter we derive a general geometric imaging model that allow the both the sensor and a thick lens to be oriented arbitrarily about their own pivots, and directly incorporate important pupil related parameters in the model. We then verify the accuracy of the model using ray-tracing in Zemax. Following the verification, we study the consequences of the model and note some of the important predictions that will ultimately allow us to build an extended depth-of-field imaging system.

### 3.1 Background

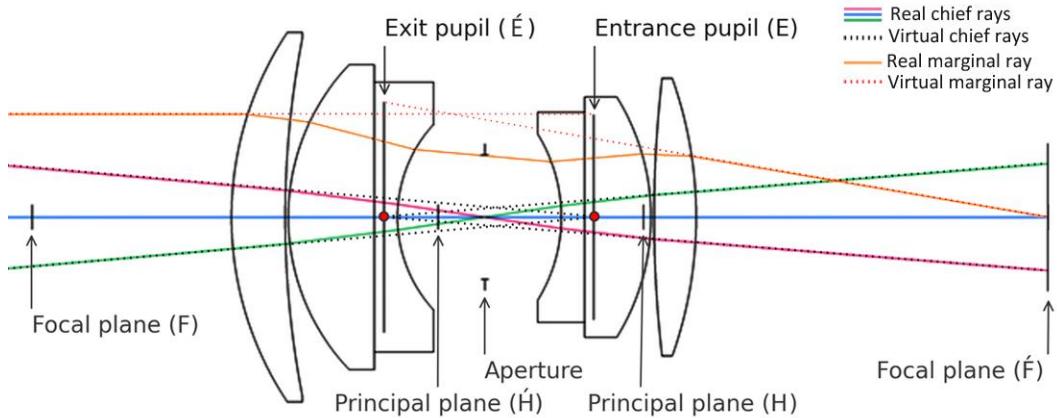
Optical imaging systems consist of several groups of elements; those elements endowed with optical power bends rays of light. The tiniest orifice in the system is called the *system aperture* or *stop*. Its interaction with the elements in the system gives rise to the pupils.

*Pupils* are the sine qua non of optical systems. They are indispensable in the design and specification of all optical systems, in both domains of *ray* and *wave optics*. The *entrance pupil* ( $E$ ) is the image of the stop seen through the elements preceding it is. The *exit pupil* ( $\bar{E}$ ) is the image of the stop seen through the elements following it is. The region preceding the entrance pupil, which includes the objects and light sources, is called the object space; and the region following the exit pupil, which includes the image plane, is called the image space. The size and position of the stop (and hence the pupils) affect image resolution, aberration, brightness, and geometry.

Rotationally symmetric lenses have an axis of symmetry—the optical axis. A ray coincident with the optical axis traverses un-deviated through the lens. Planes passing through the axis of such lenses are meridional. Rays restricted to the meridional planes are *meridional rays*. Patterns formed by the meridional rays on either side of the optical axis are mirror-reversed, exhibiting bilateral symmetry. [Figure 3.2](#) shows two types of meridional rays, traced in Zemax, that are fundamental to geometric analysis [43,72]. The *marginal ray* (MR) originates from the axial object position and skirts the edges of the aperture and pupils (virtually); the *chief ray* (CR) starts at an off-axis object point and pierces the centers of the aperture and pupils<sup>2</sup> (virtually). This pair of rays determines the location and size of the pupils, the position of the image, and the magnification. Furthermore, the bundle of chief rays from the object space converge at the center of the entrance

---

<sup>2</sup> In the presence of spherical aberrations, the chief ray goes through the center of the aperture but may not exactly go through the center of the pupils [43,73].



**Figure 3.2** Fundamental rays (contained within the meridional plane) and pupils in a Double Gauss lens for an object at infinity. The chief rays—close to the optical axis ( $0^\circ$ ,  $\pm 5^\circ$  in the object space at entrance pupil)—appear to converge at the center of entrance pupil and diverge from the center of exit pupil. The marginal ray, which is parallel to the optical axis since the object is at infinity, appear to skirt the edges of the two pupils. The red circles specify the vertices of the perspective cones (centers of the pupils). The rays were traced in Zemax.

---

pupil—thus *homocentric*—forming the vertex of the object-space perspective cone; in the image space, the bundle of chief rays diverges from the center of the exit pupil producing the vertex of the image-space perspective cone.

Imagine a film projector working backwards. Imagine the stream of light rays flowing from the illuminated portion of the scene towards a small circular hole in the projector. This pencil of rays forms a conical volume of light—the perspective cone—with its vertex at the hole and its base towards the scene. The “illuminated portion” is the angular extent of the scene visible in the image, confined by the circumferential chief rays. These extreme chief rays determine the opening angle of the cone. The “small hole” represents the entrance pupil of a camera or the pupil at the center of the iris in an eye. In the image space (behind the hole), the ray-pencil form another cone with the vertex at the center of the exit pupil. This image-space perspective cone projects the light from the scene onto the film surface or the retina in the eye. This process of image formation, known as the *central projection*, is fundamental to all imaging systems—inanimate and animate—including the camera and the eye. While the opening angle of the object-space perspective cone determines

the field-of-view, its counterpart in the image space determines the angular dimension of the image. The ratio of the pupil sizes (pupil magnification) determines the relationship between the image and object-space opening angles of the two perspective cones [74,75].

### 3.2 Notations

- *Coordinate system:* Right-handed, with the positive  $z$ -axis oriented along the direction of light travel (left to right in the plane of drawing).
- *Scalars:* written as small letters (e.g.  $x$ ). *Vectors:* small bold letters (e.g.  $\mathbf{x}$ ). *Matrices:* written as capital letters (e.g.  $M$ ).
- *Object and image space:* Non-primed quantities represent object space (e.g.  $x, \mathbf{x}$ ). Primed quantities represent image space (e.g.  $\dot{x}, \dot{\mathbf{x}}$ ).
- *Unit vectors:* represented using a hat ( $\hat{\cdot}$ ). The only exception is the direction cosine vectors which have unit norm, but are represented without the top hat (e.g.  $\mathbf{l}$ ).
- *Reference frames:* Coordinate reference frames are denoted using curly brackets (e.g.  $\{C\}$  represents the camera coordinate frame). Note that if the camera is not translated in three-dimensional space (as is the case in this thesis), a separate world frame is redundant. Therefore, in this thesis, we represent world points with respect to the camera frame  $\{C\}$ .
  - A left superscript on a variable indicates the frame of reference. For example,  ${}^0K$  indicates that the variable  $K$  is expressed respect to the world coordinate frame  $\{O\}$ . If no reference is explicitly stated, then variable is expressed with respect to the world coordinate frame (or the camera coordinate frame  $\{C\}$  if the camera coordinate frame and the world coordinate frame are the same).
- A subscript is used to associate a variable with a specific optical parameter. For example, the subscript  $e$  in variable  $d_e$  is used to represent the location of the entrance pupil, and the subscript  $\ell$  in  ${}^C R_\ell$  is used to represent the rotation matrix applied to the lens plane in the

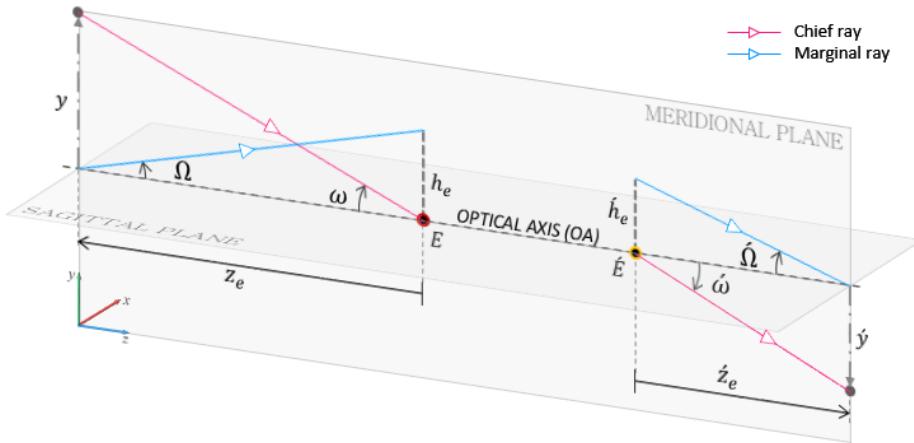
camera frame  $\{C\}$ , etc. If the camera coordinate frame is the same as the world coordinate frame, then the notation  $R_\ell$  shall be used (the superscript is dropped for convenience).

- Indices into matrices and vectors start from one (1). For example,  $\hat{\mathbf{n}}(3)$  is the third element of the unit normal vector  $\hat{\mathbf{n}}$ , and  $R_\ell(1, 2)$  is the element in the first row and second column of the matrix  $R_\ell$ . Further, a  $m \times n$  matrix  $R_\ell$  is represented in terms of its columns as  $R_\ell = [\mathbf{r}_{\ell,1} \quad \mathbf{r}_{\ell,2} \quad \dots \quad \mathbf{r}_{\ell,n}]$ , where  $\mathbf{r}_{\ell,i}$  for  $i = 1 \dots n$  are the  $n$  columns of  $R_\ell$ .

### 3.3 Relation between pupil magnification and chief ray angle

The *pupil magnification*  $m_p$  is defined as the ratio of the paraxial exit pupil diameter to the paraxial entrance pupil diameter [74,76,77].

[Figure 3.3](#) illustrates the meridional and sagittal planes associated with an arbitrarily located object of height  $y$  above the optical axis and its image of height  $\dot{y}$  in a typical optical system. The figure also shows the chief ray from the object's edge further from the optical axis, the marginal ray from the axial point in the object, and the two pupils contained in the meridional plane. The schematic, although simple, is quite general as a (meridional) plane always exist for a given object point irrespective of its position in the three-dimensional space, if the lens is rotationally symmetric.



**Figure 3.3** Schematic of chief and marginal rays. The ratio of the tangents of the chief ray angles in the object space to the image space yields the pupil magnification.

Let the angles between the chief ray and the optical axis (called the *ray-angle*) in the object and image space be  $\omega$  and  $\dot{\omega}$  respectively. Also, let the angles produced by the marginal ray with the optical axis in the object and image space be  $\Omega$  and  $\dot{\Omega}$  respectively. Then, we can obtain the relation between the chief-ray ray-angles— $\omega$  and  $\dot{\omega}$ —and the pupil magnification  $m_p$  as follows:

From the [Figure 3.3](#) we obtain,

$$\begin{aligned}\tan(\Omega) &= \frac{h_e}{z_e}, & \tan(\omega) &= \frac{y}{z_e}, \\ \tan(\dot{\Omega}) &= \frac{h_e}{\dot{z}_e}, & \tan(\dot{\omega}) &= \frac{\dot{y}}{\dot{z}_e}.\end{aligned}\tag{3.1}$$

Eliminating  $z_e$  and  $\dot{z}_e$  after dividing  $\tan(\dot{\omega})$  by  $\tan(\omega)$ , we have

$$\frac{\tan(\omega)}{\tan(\dot{\omega})} = \frac{h_e}{h_e} \frac{y \tan(\Omega)}{\dot{y} \tan(\dot{\Omega})} \approx \frac{h_e}{h_e} \frac{y \Omega}{\dot{y} \dot{\Omega}}.\tag{3.2}$$

A common observation in imaging is that we can increase the transverse magnification ( $\dot{y}/y$ ) by increasing the lens-to-image-plane distance while correspondingly decreasing the lens-to-object-plane distance to maintain focus on the object. However, increasing (decreasing) the image plane distance proportionally decreases (increases) the marginal ray angle  $\dot{\Omega}$  (see [Figure 3.3](#)). Consequently, the angular magnification ( $\dot{\Omega}/\Omega$ ) decreases with increase in lens-to-image-plane distance. Therefore, a large transverse magnification is associated with a correspondingly small angular magnification. This result follows from a more general theory called the *Lagrange invariant* property [74] of the two rays (the chief ray and the marginal ray) when applied between conjugate locations. As per the invariant property, the product of the transverse magnification and the angular magnification equals to one, i.e.,  $\left(\frac{\dot{y}}{y}\right) \left(\frac{\dot{\Omega}}{\Omega}\right) = 1$  or  $y\Omega = \dot{y}\dot{\Omega}$ . Cancelling the corresponding terms in Eq. (3.2) yields the relationship between the pupil magnification and the object and image space chief ray angles as:

$$\boxed{\frac{\tan(\omega)}{\tan(\dot{\omega})} = m_p} \quad (3.3)$$

where,

$m_p$	Pupil magnification.
$\omega$	Angle between the chief ray and the optical axis in the object (input) space.
$\dot{\omega}$	Angle between the chief ray and the optical axis in the image (output) space.

The above relation (Eq. (3.3)) has been previously derived in [77] using a different formulation.

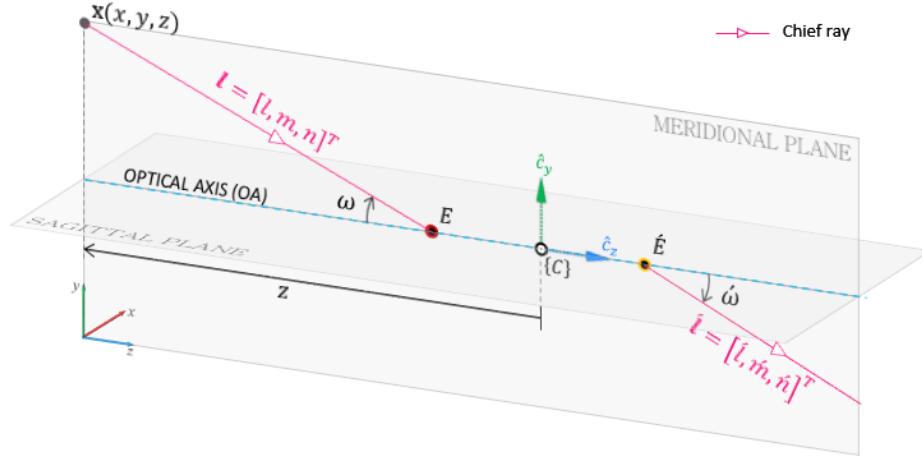
For a given optical system with fixed focal length, the pupil magnification  $m_p$  is constant. This constancy of the ratio of the tangents of the chief ray angles for varying object (and image) heights is a necessary and sufficient condition for distortion-free imaging known as the *Airy's Tangent-Condition* [73,77]. Eq. (3.3) also suggests that when  $m_p = 1$  the perspective cones in the object- and image-space are symmetric. In the following section, we will use Eq. (3.3) to derive the relationship between the direction cosines of the object-space (input) chief rays and direction cosines of the image-space (output) chief rays.

### 3.4 Transfer of chief ray's direction cosines between the pupils

The direction cosines, a unit vector of cardinality three, specify the direction of a ray. Its elements are the cosines of the angles the ray makes with the three coordinate axes. In other words, the elements of the direction cosine vector are the projections of the unit vector in the direction of the ray on the  $x$ -,  $y$ -, and  $z$ -axes. In the absence of aberrations, the chief ray starts from an object point, passing through the center of the entrance pupil (virtually), the aperture stop, exit pupil (virtually), and ends at the image point. Suppose we know the direction cosine of the chief ray in the object space (between the object point and the entrance pupil), what is the direction cosine of the chief

ray in the image space (between the exit pupil and image point)? Furthermore, what is the relation between the input and output chief ray's direction cosines if the lens is swiveled about a pivot point along the optical axis?

We begin by solving a specific problem of the *transfer* of the direction cosines between the pupils in which the optical axis coincides with the  $z$ -axis of the camera frame  $\{C\}$ . The configuration of this specific problem is show in [Figure 3.4](#). Subsequently, we will deduce the general *transfer* expression in which the optical axis is free to swivel about the origin of  $\{C\}$ . Let  $\mathbf{l} = [l, m, n]^T$  be the direction cosine of the chief ray from an object point  $\mathbf{x}$  to the center of the entrance pupil, and let  $\hat{\mathbf{l}} = [\hat{l}, \hat{m}, \hat{n}]^T$  be the corresponding direction cosine of the chief ray from the exit pupil to the image point. The parameters  $\mathbf{x}$ ,  $\mathbf{l}$ , and  $\hat{\mathbf{l}}$  are specified with respect to frame  $\{C\}$ .



**Figure 3.4** Specific problem—optical axis coincides with reference frame's  $z$ -axis. If  $\theta$  and  $\hat{\theta}$  are the angles of the chief ray with the optical axis in the object and image space respectively, then  $\omega = \theta = \cos^{-1}(n)$  and  $\hat{\omega} = \hat{\theta} = \cos^{-1}(\hat{n})$ .

If  $\theta$  and  $\phi$  are the zenith and azimuthal angles of the chief ray in the object space, and  $\hat{\theta}$  and  $\hat{\phi}$  the corresponding angles in the image space, then the direction cosines, in  $\{C\}$ , are:

$$\begin{aligned} l &= \cos(\phi) \sin(\theta) & \hat{l} &= \cos(\phi) \sin(\hat{\theta}) \\ m &= \sin(\phi) \sin(\theta) & \hat{m} &= \sin(\phi) \sin(\hat{\theta}) \\ n &= \cos(\theta) & \hat{n} &= \cos(\hat{\theta}) \end{aligned} \quad (3.4)$$

Since the optical axis is aligned with the  $z$ -axis, we have  $\omega = \theta$  and  $\dot{\omega} = \hat{\theta}$ . Substituting the expressions for  $\sin(\theta)$ ,  $\cos(\theta)$ ,  $\sin(\hat{\theta})$  and  $\cos(\hat{\theta})$  from Eq. (3.4) into Eq. (3.3) we obtain:

$$\begin{aligned} \frac{l n}{\hat{l} \hat{n}} \cos(\phi) &= m_p \cos(\phi) \\ \text{and} \\ \frac{m n}{\hat{m} \hat{n}} \sin(\phi) &= m_p \sin(\phi) \end{aligned} \quad (3.5)$$

Further, since the input and output chief rays are confined to the same meridional plane [43,72],  $\hat{\phi} = \phi$ , yielding  $\hat{l}$  and  $\hat{m}$  in terms of  $l$  and  $m$ , the ratios of  $\hat{n}$  to  $n$ , and  $m_p$ :

$$\begin{aligned} \hat{l} &= \frac{1}{m_p} \frac{\hat{n}}{n} l , \\ \text{and} \\ \hat{m} &= \frac{1}{m_p} \frac{\hat{n}}{n} m . \end{aligned} \quad (3.6)$$

From (3.3) we have

$$m_p = \frac{\tan(\theta)}{\tan(\hat{\theta})} = \frac{\sin(\theta) \cos(\hat{\theta})}{\sin(\hat{\theta}) \cos(\theta)} = \sqrt{\frac{1-n^2}{1-\hat{n}^2}} \times \frac{\hat{n}}{n} , \quad (3.7)$$

which after simplification yields  $\hat{n}$  in terms of the pupil magnification  $m_p$  and input  $n$  as

$$\dot{n} = \pm \frac{m_p}{\sqrt{1 + (m_p^2 - 1)n^2}} n . \quad (3.8)$$

Combining Eqs. (3.6) and (3.8), we obtain the expression for output direction cosine of the chief ray in terms of its input direction cosines and the pupil magnification as:

$$\begin{bmatrix} \dot{l} \\ \dot{m} \\ \dot{n} \end{bmatrix} = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n^2}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & m_p \end{bmatrix} \begin{bmatrix} l \\ m \\ n \end{bmatrix} , \quad (3.9)$$

which we can write compactly as:

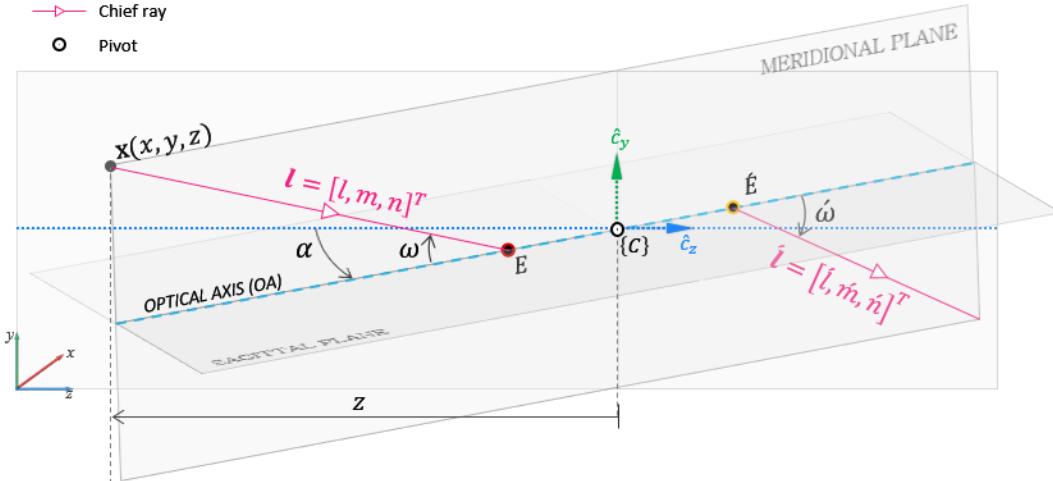
$$\dot{\mathbf{l}} = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n^2}} M_p \mathbf{l} . \quad (3.10)$$

Our objective is to derive the expression for the transfer of the chief ray's direction cosines from entrance pupil to exit pupil for arbitrary orientation of the optical axis as shown in [Figure 3.5](#). Although a formal derivation is provided in [Appendix A.1](#), we can readily infer the general expression for the *transfer* from Eq. (3.10). Suppose we swivel the optical axis about the origin of the camera frame  $\{C\}$ . This rotation can be described by the matrix  $R_\ell \in \mathbb{R}^{3 \times 3}$ . As before, we designate the ray from the object point  $\mathbf{x}$  to the (new position of the) center of the entrance pupil as the chief ray. Let us also suppose that we have another coordinate frame,  $\{\mathcal{L}\}$ , sharing its origin with  $\{C\}$  and its  $z$ -axis coincident with the optical axis. If  $\mathbf{l}$  be the direction cosine of the chief ray from the object point in the frame  $\{C\}$ , then the direction cosine in the frame  $\{\mathcal{L}\}$  is  $R_\ell^T \mathbf{l}$  and the third element of the direction cosine is  $\mathbf{r}_{\ell,3}^T \mathbf{l}$ , where  $\mathbf{r}_{\ell,3}$  is the third column of  $R_\ell$ . Representing  $n_R = \mathbf{r}_{\ell,3}^T \mathbf{l}$ , the direction cosine of the chief ray emerging from the exit pupil is obtained by substituting  $\mathbf{r}_{\ell,3}^T \mathbf{l}$  for  $\mathbf{l}$  and  $n_R$  for  $n$  in Eq. (3.10):

$$\hat{\mathbf{l}} = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} M_p R_\ell^T \mathbf{l}$$

The above expression represents the output direction cosine in the coordinate frame  $\{\mathcal{L}\}$ . In order to transform the output direction cosine from the coordinate frame  $\{\mathcal{L}\}$  to the camera frame  $\{\mathcal{C}\}$  we need to multiply the direction cosine vector by  $R_\ell$  to obtain:

$$\hat{\mathbf{l}} = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} R_\ell M_p R_\ell^T \mathbf{l} \quad (3.11)$$



**Figure 3.5** Configuration of the general problem—optical axis pivots freely about the origin of camera frame  $\{\mathcal{C}\}$ .

---

The positive or negative sign of the direction cosine determines the forward or backward direction of light-travel along a rectilinear path. Under the assumptions of isotropy and homogeneity, the only condition under which a ray of light emerges in an antipodal path from an interface is if it encounters a mirror surface *normally*. This condition does not arise within the context of our problem. Therefore, without any loss of generality, we can drop the negative sign

in Eq. (3.11); accordingly, the output direction cosines assume the sign of the corresponding input direction cosines. Therefore, we obtain the general expression for the direction cosines of the chief ray in the image space as:

$$\boxed{\boldsymbol{l} = \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} R_\ell M_p R_\ell^T \boldsymbol{l}}$$
 (3.12)

where,

$n_R$	The third element of the input direction cosine vector following rotation. $n_R = \mathbf{r}_{\ell,3}^T \boldsymbol{l}$
$\boldsymbol{l}$	Input or object space chief ray's direction cosine vector from the object point $\mathbf{x}$ to the entrance pupil.
$\tilde{\boldsymbol{l}}$	Output or image space chief ray's direction cosine vector emerging from the exit pupil.
$M_p$	Equal to $\text{diag}(1, 1, m_p)$ , where $m_p$ is the pupil magnification.
$R_\ell$	Rotation matrix used to describe the orientation of the lens plane.
$\mathbf{r}_{\ell,3}$	The third column of $R_\ell$ .

Note that Eq. (3.12) only describes the output chief ray's direction cosines—a free vector. The exact output chief ray is obtained from the knowledge of the direction cosine and the location of the exit pupil in the appropriate reference frame.

Although it is not obvious from the expression Eq. (3.12), we expect  $\tilde{\boldsymbol{l}}$  to have unit magnitude. We present a proof in [Appendix A.2](#) that shows the  $\ell^2$ -Norm (magnitude) of  $\tilde{\boldsymbol{l}}$  is indeed equal to one, and  $\{1 + (m_p^2 - 1)n_R^2\}^{-1/2}$  is the normalizing term.

Furthermore, we can draw the following inferences about  $\tilde{\boldsymbol{l}}$  from the Eq. (3.12):

1. If the pupil magnification,  $m_p = 1$ , then  $\tilde{\boldsymbol{l}} = \boldsymbol{l}$ , which implies that the opening angles of the image and object space perspective cones are equal, irrespective of the orientation of

the optical axis. Then, the lens is symmetric about a plane perpendicular to the optical axis (in addition to the rotational symmetry about the optical axis). It must not come as a surprise that symmetric lenses are can be reversed without affecting any optical system properties [76].

2. If we let  $Q = R_\ell$ , such that  $A = QM_pQ^T$ , then we can write  $\mathbf{l} \cong kA\mathbf{l}$ , where  $k$  is the scalar normalization term. Furthermore, as  $M_p$  is a diagonal matrix, and  $Q$  is orthonormal, we can immediately recognize the form  $A = QM_pQ^T$  as the Eigen value decomposition of a symmetric matrix  $A$ , with  $\mathbf{q}_i = \mathbf{r}_{\ell,i}$  for  $i = 1, 2, 3$ —the columns of  $Q$ —as the eigenvectors and  $\lambda = \{1, 1, m_p\}$  the corresponding eigenvalues. i.e., as  $M_p$  is a diagonal matrix,

$$\begin{aligned} A &= QM_pQ^T \\ &= \sum_{i=1}^3 M_p(i,i)\mathbf{q}_i\mathbf{q}_i^T , \end{aligned}$$

and

$$\begin{aligned} \mathbf{l} &= k \sum_{i=1}^3 M_p(i,i)\mathbf{q}_i\mathbf{q}_i^T \mathbf{l} \\ &= k(\mathbf{q}_1^T \mathbf{l})\mathbf{q}_1 + k(\mathbf{q}_2^T \mathbf{l})\mathbf{q}_2 + k m_p(\mathbf{q}_3^T \mathbf{l})\mathbf{q}_3 . \end{aligned} \quad (3.13)$$

In Eq. (3.13) the terms  $(\mathbf{q}_i^T \mathbf{l})$  for  $i = 1, 2, 3$  are the projections of the input direction cosine along the eigenvectors  $\mathbf{q}_i = \mathbf{r}_{\ell,i}$ . Also,  $\mathbf{q}_3 = \mathbf{r}_{\ell,3}$ , the third column of the rotation matrix  $R_\ell$ , is the direction of the optical axis.

### 3.5 Image formation for arbitrary orientation of the lens and image plane

Geometric imaging is a mapping (*bijective* in projective space) between points in the three-dimensional world space to corresponding points on a mathematical surface that we call the *image*. Here we aim to study the nature of this mapping on a planar surface—the image plane—for

arbitrary orientations of the lens and image planes. To that effect, we will use the knowledge of the transfer of direction cosines of the chief ray derived previously.

An extended object emanates a multitude of chief rays that reach the image space through the lens elements and the stop. The locus of points formed by the intersection of these rays with the image plane constitutes the *projection* of the object in the image plane [78,79]. Further, we generally identify the projection of the object point as an “image” when the pencil of rays from the object point geometrically converge at a single point in the image space.

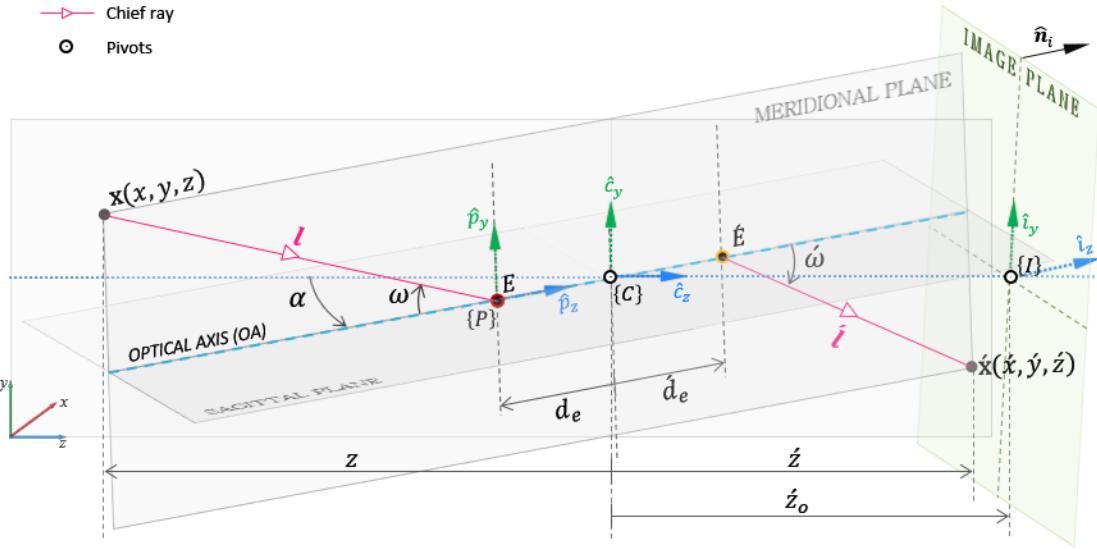
For simplicity, we assume that the lens is unencumbered by radial distortions and optical aberrations. [Figure 3.6](#) represents a schematic of the problem in which we have introduced an image plane whose orientation is described by its surface normal  $\hat{\mathbf{n}}_i$ . Two local frames are also introduced: the frame  $\{P\}$  is attached to the optical axis with its origin at entrance pupil, and the frame  $\{I\}$  attached to the image plane with its origin at the image plane pivot. The image plane is free to tilt or swing about its local x-axis or y-axis respectively at the image plane pivot.

Let the exit pupil ( $\tilde{E}$ ) be located  $\tilde{d}_e$  units from the pivot point along the optical axis. Following the rotation of the optical axis, by applying the matrix  $R_\ell \in \mathbb{R}^{3 \times 3}$ , the position of the exit pupil in camera frame  $\{C\}$  is given as  $R_\ell[0,0,\tilde{d}_e]^T = \tilde{d}_e \mathbf{r}_{\ell,3}$ .

We can represent the chief ray emerging from the exit pupil with direction cosine  $\mathbf{l}$  by the parametric equation:

$$\dot{\mathbf{x}}(\lambda) = \tilde{d}_e \mathbf{r}_{\ell,3} + \frac{\lambda}{\sqrt{1 + (m_p^2 - 1)n_R^2}} R_\ell M_p R_\ell^T \mathbf{l}, \quad (3.14)$$

where  $\dot{\mathbf{x}}(\lambda)$  represent points along the output chief ray in  $\{C\}$ . The first term on the right hand side of Eq. (3.14) is the initial position of the ray (at the center of  $\tilde{E}$ ) and  $\lambda$  is a real number that determines the length of the ray.



**Figure 3.6** Schematic of geometric image formation.  $\hat{\mathbf{x}}$  is the *central projection* of the object point  $\mathbf{x}$  on image plane. The optical axis and image plane are free to swivel about the origins of coordinate frames  $\{C\}$  and  $\{I\}$  respectively.

We write the equation of the image plane in Hessian normal form as:

$$\hat{\mathbf{n}}_i^T \xi = \dot{z}_{o\perp}, \quad (3.15)$$

where  $\hat{\mathbf{n}}_i$  is the unit normal to the image plane,  $\dot{z}_{o\perp}$  is the perpendicular distance of the plane from the origin of frame  $\{C\}$ , and  $\xi$  is a point on the image plane.

We obtain the expression for  $\lambda$  (in Eq. (3.14)) for which the ray intersects the image plane by equating  $\xi$  to  $\hat{\mathbf{x}}$ , multiplying Eq. (3.14) by  $\hat{\mathbf{n}}_i^T$ , and rearranging the terms:

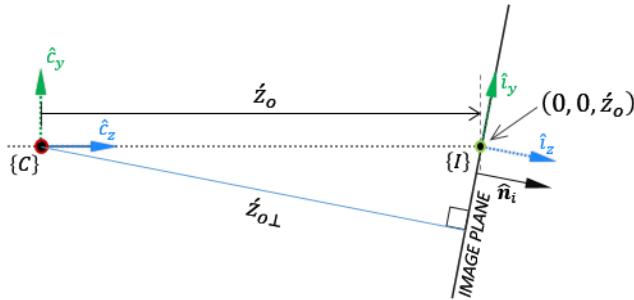
$$\begin{aligned} \hat{\mathbf{n}}_i^T \hat{\mathbf{x}} &= \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3} + \frac{\lambda}{\sqrt{1 + (m_p^2 - 1)n_R^2}} \hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T \mathbf{l}, \\ \lambda &= \frac{(\dot{z}_{o\perp} - \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3}) \sqrt{1 + (m_p^2 - 1)n_R^2}}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T \mathbf{l}}. \end{aligned} \quad (3.16)$$

Substituting Eq. (3.16) into Eq. (3.14) we get:

$$\dot{\mathbf{x}} = \dot{d}_e \mathbf{r}_{\ell,3} + \frac{(\dot{z}_{o\perp} - \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T \mathbf{l}} R_\ell M_p R_\ell^T \mathbf{l} \quad (3.17)$$

Now we proceed to find the expression for the perpendicular distance  $\dot{z}_{o\perp}$  in terms of the known parameters. The origin of the image plane's local reference frame,  $\{I\}$ , is located at the intersection of the z-axis of camera frame  $\{C\}$  with the image plane (see [Figure 3.7](#)). Given the location of the image plane's pivot,  $(0,0,\dot{z}_o)$ , we can describe the orientation of the image plane using the surface normal. The image plane's surface normal is obtained by applying the rotation matrix  $R_i$  to the unit vector  $[0,0,1]^T$ :

$$\hat{\mathbf{n}}_i = R_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (3.18)$$



**Figure 3.7** Schematic of the image plane. The image plane having surface normal  $\hat{\mathbf{n}}_i$  is located at  $\dot{z}_o$  units from the origin of camera frame  $\{C\}$  along the z-axis that intersects the plane at  $(0,0,\dot{z}_o)$ .  $\dot{z}_{o\perp}$  is the perpendicular distance from the origin to the plane. The local image coordinate frame with its origin at the intersection of the image plane and z-axis of the camera frame is represented by  $\{I\}$ .

---

The equation of the image plane as represented by Eq. (3.15) is

$$\hat{\mathbf{n}}_i^T \xi = \dot{z}_{o\perp}.$$

Since  $\xi = [0, 0, \dot{z}_o]^T$  is a point on the plane, we obtain  $\dot{z}_{o\perp}$  as:

$$\dot{z}_{o\perp} = \hat{\mathbf{n}}_i^T \begin{bmatrix} 0 \\ 0 \\ \dot{z}_o \end{bmatrix} = \hat{\mathbf{n}}_i(3) \dot{z}_o . \quad (3.19)$$

Substituting  $\dot{z}_{o\perp}$  from Eq. (3.19) into Eq. (3.17) yields the expression for the point of intersection of the chief ray with the image plane in terms of the input direction cosines as

$$\dot{\mathbf{x}} = \dot{d}_e \mathbf{r}_{\ell,3} + \frac{(\hat{\mathbf{n}}_i(3) \dot{z}_o - \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T \mathbf{l}} R_\ell M_p R_\ell^T \mathbf{l}. \quad (3.20)$$

Let the entrance pupil be located at a distance  $d_e$  from the pivot point along the optical axis in the camera frame  $\{C\}$ . Then, similar to the description of the exit pupil, the location of the entrance pupil in  $\{C\}$  is given as:

$$\mathbf{x}_e = R_\ell [0, 0, d_e]^T = d_e \mathbf{r}_{\ell,3} . \quad (3.21)$$

Further, we express the direction cosine components in terms of the Cartesian coordinates of the object point and entrance pupil in the camera frame  $\{C\}$  as:

$$\begin{aligned} l &= \frac{x_e - x}{\sqrt{(x_e - x)^2 + (y_e - y)^2 + (z_e - z)^2}} = \frac{d_e R_\ell(1,3) - x}{\|\mathbf{x}_e - \mathbf{x}\|} , \\ m &= \frac{y_e - y}{\sqrt{(x_e - x)^2 + (y_e - y)^2 + (z_e - z)^2}} = \frac{d_e R_\ell(2,3) - y}{\|\mathbf{x}_e - \mathbf{x}\|} , \\ n &= \frac{z_e - z}{\sqrt{(x_e - x)^2 + (y_e - y)^2 + (z_e - z)^2}} = \frac{d_e R_\ell(3,3) - z}{\|\mathbf{x}_e - \mathbf{x}\|} , \end{aligned} \quad (3.22)$$

which we write compactly as:

$$\mathbf{l} = \frac{-(\mathbf{x} - d_e \mathbf{r}_{\ell,3})}{\|\mathbf{x}_e - \mathbf{x}\|} . \quad (3.23)$$

Substituting Eqs. (3.23) and (3.21) into Eq. (3.20), we obtain a general relationship between the object point  $\mathbf{x}$  and its corresponding image point  $\dot{\mathbf{x}}$  as:

$$\dot{\mathbf{x}} = \dot{d}_e \mathbf{r}_{\ell,3} + \frac{(\hat{\mathbf{n}}_i(3)\dot{z}_o - \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T (\mathbf{x} - d_e \mathbf{r}_{\ell,3})} R_\ell M_p R_\ell^T (\mathbf{x} - d_e \mathbf{r}_{\ell,3}) \quad (3.24)$$

Eq. (3.24) represents the image point  $\dot{\mathbf{x}}$  in the camera frame. Once an image—a two dimensional representation of the scene—has been formed, we specify positions and dimensions in the image independent of the position and orientation of the sensor and lenses (e.g. in terms of pixels in a digital image). We can transform the image coordinates in the camera frame  $\{C\}$  represented by Eq. (3.24) to the image frame  $\{I\}$  by observing that the origin of  $\{I\}$  is displaced from  $\{C\}$  by the translation vector  $\mathbf{t}_i = [0, 0, \dot{z}_o]^T$ , and the standard basis vectors of frame  $\{I\}$  are rotated by  $R_i \in \mathbb{R}^{3 \times 3}$ . Consequently, a point  ${}^I\dot{\mathbf{x}}$  in  $\{I\}$  relative to  $\{C\}$  may be expressed as (see Eq. 2.53 in [80]):

$${}^C\dot{\mathbf{x}} = R_i {}^I\dot{\mathbf{x}} + \mathbf{t}_i. \quad (3.25)$$

Therefore, we can write the expression for the image point coordinates in the image plane's reference frame as:

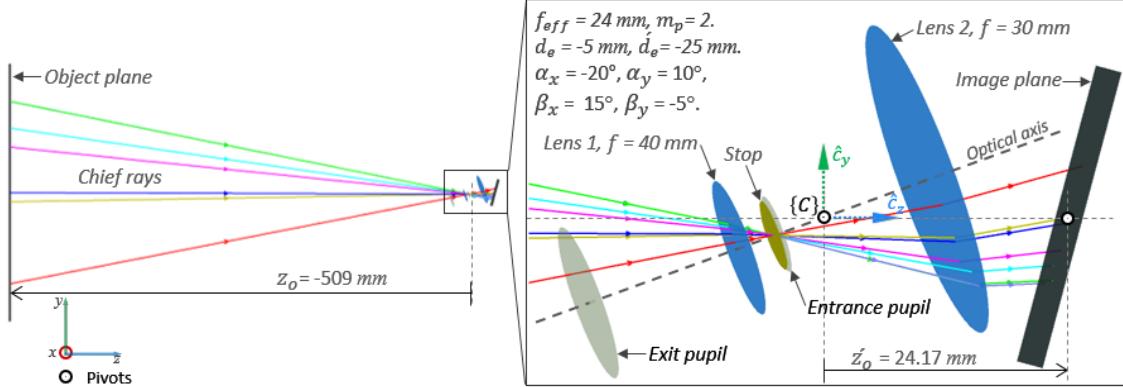
$${}^I\dot{\mathbf{x}} = R_i^T ({}^C\dot{\mathbf{x}} - \mathbf{t}_i) \quad (3.26)$$

Substituting  ${}^C\dot{\mathbf{x}}$  from Eq. (3.24) into Eq. (3.26) we obtain the expression of the two-dimensional image point in the image frame as:

$${}^I\dot{\mathbf{x}} = R_i^T (\dot{d}_e \mathbf{r}_{\ell,3} - \mathbf{t}_i) + \frac{(\hat{\mathbf{n}}_i(3)\dot{z}_o - \dot{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T ({}^C\mathbf{x} - d_e \mathbf{r}_{\ell,3})} R_i^T R_\ell M_p R_\ell^T ({}^C\mathbf{x} - d_e \mathbf{r}_{\ell,3}) \quad (3.27)$$

Where,

$c_x$	3D Cartesian coordinates (in physical units) of the world point in camera frame $\{C\}$ . In the adopted coordinate convention, the numerical value of z-component of $\mathbf{x}$ is negative.
${}^I \dot{\mathbf{x}}$	2D Cartesian coordinates (in physical units) of the image point in the image frame $\{I\}$ . Note that Eq. (3.27) produces a $3 \times 1$ vector with the third element (i.e. the ${}^I \dot{z}$ component) identically equal to zero.
$M_p$	Equal to $\text{diag}(1, 1, m_p)$ , where $m_p$ is the pupil magnification.
$d_e$	Location of the entrance pupil from the pivot (origin of $\{C\}$ ) along the optical axis. This scalar quantity physical units, usually in millimeters.
$\dot{d}_e$	Location of the exit pupil from the pivot (origin of $\{C\}$ ) along the optical axis.
$\dot{z}_o$	Location of the image plane from the pivot (origin of $\{C\}$ ) along the z-axis of $\{C\}$ .
$R_\ell$	Rotation matrix used to describe the orientation of the lens plane.
$r_{\ell,3}$	The third column of $R_\ell$ .
$R_i$	Rotation matrix used to describe the orientation of the image plane.
$\hat{\mathbf{n}}_i$	The image plane normal. $\hat{\mathbf{n}}_i(3)$ denotes the z-component of the normal.
$t_i$	Origin of the image frame with respect to camera frame; $\mathbf{t}_i = [0, 0, \dot{z}_o]^T$ .



**Figure 3.8** Ray tracing for verifying Eq. (3.27). Chief rays traced from a grid of points in the object plane through an ideal lens tilted about a point  $d_e = -5 \text{ mm}$  away from the entrance pupil along the optical axis to the tilted image plane.

### 3.6 Verification of imaging equation in Zemax

We derived Eq. (3.27), which relates a three-dimensional object point to its projection in the two-dimensional image plane of a Scheimpflug camera, analytically. Now we verify the accuracy of the relationship by comparing the numerically computed values of image points (intersection of chief ray with the image plane) using Eq. (3.27) with corresponding image points obtained by tracing chief rays from a grid of points belonging to the object plane. [Figure 3.8](#) is a layout plot of the optical system modeled in Zemax showing (1) an object plane, (2) an ideal lens made from two paraxial surfaces and pivoted about a point away from the entrance pupil ( $d_e = -5\text{ mm}$ ), and (3) an image plane pivoted about the image plane pivot along the  $z$ -axis. We can arbitrarily assign any rotation angle to both the lens and image planes (pivot at their local frame's origin) with respect to both  $x$ - and  $y$ -axis. The orientation of both planes is represented using *intrinsic* rotations matrices (composed of elemental rotations first about the  $x$ -axis followed by rotation about the new  $y$ -axis). Symbols  $\alpha_x$  and  $\alpha_y$  represent the angles of rotation of the lens plane about the  $x$ - and  $y$ -axes while  $\beta_x$  and  $\beta_y$  represent the angles of rotation of the image plane about the  $x$ - and  $y$ -axes. The results of the simulation are tabulated in [Table 3.1](#), which shows the set of object points, the numerically computed image points, the ray traced image points, and the absolute difference between the numerically computed and ray traced image points. We observe that the numerically computed and ray traced values of the image points are very close; the small difference in their values can be attributed to the error associated with floating point operations. This comparison demonstrates that the analytically derived expression (Eq. (3.27)) representing geometric relationship between a three-dimensional object point and its image point in the absence of optical aberrations is accurate.

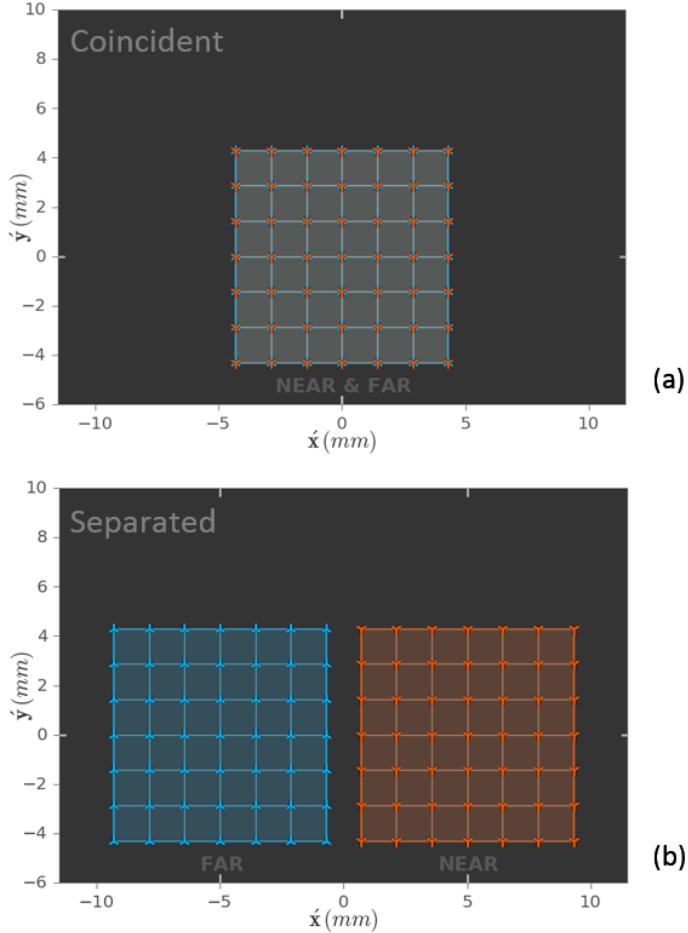
**Table 3.1** Comparison of numerically computed image points with ray traced (in Zemax) image points for the optical system shown in [Figure 3.8](#).

World point $c\mathbf{x}(x, y, z)$	Computed image point ${}^I\hat{\mathbf{x}}(\hat{x}, \hat{y}, \hat{z})$	Ray-traced image point ${}^I\hat{\mathbf{x}}_{rt}(\hat{x}, \hat{y}, \hat{z})$	Absolute difference $ {}^I\hat{\mathbf{x}} - {}^I\hat{\mathbf{x}}_{rt} $
(0.0, 0.0, -509.0)	(-0.3108, -0.6291, 0.0)	(-0.3108, -0.6291, 0.0)	(1.8e-09, 3.1e-09, 7.5e-15)
(10.0, -10.0, -509.0)	(-0.8003, -0.0863, 0.0)	(-0.8003, -0.0863, 0.0)	(2.1e-09, 2.7e-09, 3.0e-15)
(-50.0, 50.0, -509.0)	(2.1291, -3.3352, 0.0)	(2.1291, -3.3352, 0.0)	(1.2e-09, 3.2e-09, 2.9e-15)
(70.71, 70.71, -509.0)	(-4.2013, -5.0221, 0.0)	(-4.2013, -5.0221, 0.0)	(2.6e-09, 5.1e-09, 4.7e-15)
(100.0, 0.0, -509.0)	(-5.5251, -1.0101, 0.0)	(-5.5251, -1.0101, 0.0)	(1.3e-09, 8.4e-09, 3.1e-15)
(0.0, 100.0, -509.0)	(-0.6031, -6.4387, 0.0)	(-0.6031, -6.4387, 0.0)	(2.2e-09, 4.0e-09, 2.2e-16)
(100.0, 100.0, -509.0)	(-5.8238, -6.8542, 0.0)	(-5.8238, -6.8542, 0.0)	(5.6e-10, 2.5e-10, 2.2e-15)

### 3.7 Geometric properties of images under lens and image plane rotation

Following the verification of Eq. (3.27), we use the expression to qualitatively study the effects of lens and sensor rotations on the geometric properties of the image. We also investigate the effects of pupil magnification,  $m_p$ , and location of the lens pivot on the nature of the geometric distortions. [Figures 3.10 – 3.15](#) show the type of distortions in images—of two planes in the object space—for several lens and sensor orientations. In all these figures, the basic setup is similar to that shown in [Figure 3.8](#) except that the object space consists of two planes—a near plane and a far plane. The near plane is a square of 88.15 mm on each side, and the far plane is a square of 178.3 mm on each side placed at twice the distance of the near plane from the entrance pupil. The exact distance of the near plane (and consequently the far plane) from the lens vary depending upon the pupil magnification, such that the images of the two planes are 4.5 mm on each side on the sensor. Also, since the  $z$ -axis of the camera frame passes through the center of both object planes, the two images are coincident in the frontoparallel configuration (i.e. when the object planes are parallel to lens and image planes).

The object points consist of  $7 \times 7$  square grids on each of the object planes. The corresponding “image points” are the points of intersection of the chief-rays (emanating from the object points) with the image plane ([Figure 3.9 \(a\)](#)). The orange “Y” markers represent the group of image points from the near object plane. The blue “inverted Y” markers represent the image points from the far object plane. Note that in frontoparallel configuration the two images of the two object planes coincide; however, for the sake of visual clarity, we separate the two set of image points horizontally by 5 mm on either side from the center ([Figure 3.9 \(b\)](#)). In [Figures 3.10 – 3.15](#) while lighter shaded (orange and blue) markers are used to represent the points in frontoparallel configuration, darker shaded markers of either color represent the image points following the rotation of the sensor or lens. Rotation of either the lens or the sensor induces a geometric distortion of the image field in which the points across the image field translates by different amounts and directions. These translations are shown by the gray-to-white arrows between the original and shifted positions (drawn if the magnitude of the shift is greater than a certain threshold). The white level of the arrows specifies the normalized magnitude of translation—brighter indicates relatively larger translation. The figures also display information about the standard deviation (SD) of the arrow lengths. This statistic gives a sense of the non-uniform translation of the image points across the image field. If all image points shift by the same amount, then the standard deviation will be zero. A larger value of the standard deviation indicates greater diversity in shifts, and hence greater distortion. In addition to the standard deviation, we also measure how much the centroid of the set of points from the two images shifts. The translation of the centroid gives a sense of how the total image field “appear” to shift. However, we must note that we consider an image field to have *translated* only when *all* image points have at least a fixed minimum amount of shift in the direction of translation. Note that in all cases shown here, the *image points* were not determined using a “best focus” criterion, but rather by the point of intersection of the chief rays with the image plane. However, this definition of the *image* adopted for the current discussion does not limit the study of



**Figure 3.9** “Image points” corresponding to two object planes—a far plane twice the size of the near plane. (a) The image points are coincident. (b) The coincident image points are separated laterally for the purpose of our investigation.

---

geometric properties, such as the kind of transformations induced by the rotations of the sensor and lens planes.

### 3.7.1 Properties of image field induced by sensor rotation ( $\alpha_x, \alpha_y = 0, \beta_x, \beta_y \in \mathbb{R}$ )

[Figure 3.10](#) shows the images of the two object planes obtained under sensor rotation for three different values of pupil magnification. Studying the figures, we observe that:

1. The image plane is no longer frontoparallel with the object plane. As a result, the transverse magnification varies across the image field. Therefore, the image points undergo a field dependent and asymmetrical geometric distortion.

2. Since the location of both the entrance pupil and the exit pupil remain fixed, the on-axis image point continues to remain on-axis subsequent to the rotations. Therefore, we can conclude that there is no translation of the image field following sensor rotation.
3. The amount of perspective distortion is directly proportional to the pupil magnification.
4. However, as shown in [Figure 3.11](#), the distortion is independent of the object distance. Therefore, rotating the sensor plane does not introduce parallax between images that are obtained under varying orientations of the sensor. Consequently, if we capture multiple images under several rotations of the sensor, the inter-image homography—mapping between corresponding points of two images—is simply a perspective mapping of the following form:

$$H = \begin{bmatrix} a & b & 0 \\ d & e & 0 \\ g & h & 1 \end{bmatrix} \quad (3.28)$$

### ***3.7.2 Properties of image field induced by lens rotation away from center of the entrance pupil***

**$(\alpha_x, \alpha_y \in \mathbb{R}, \beta_x, \beta_y = 0; d_e \neq 0)$**

The influence of object distance and pupil magnification on the image shape obtained under lens rotation away from the entrance pupil is depicted in [Figure 3.12](#). The emergence of parallax in the images is highlighted in [Figure 3.13](#), in which we have plotted the top, middle and bottom rows of the image points for the three cases of pupil magnifications. We observe that:

1. The dominant effect of rotating the lens about a point along the optical axis is, in general, a *non-uniform shift* of the image field that depends on both the pupil magnification and object distance.
2. Since points in the image field undergo non-uniform translation, the standard deviation of the translation vector length is non-zero. Also, because the amount of shift of the image

field is dependent on the object distance, the value of the standard deviation of the translation vector lengths is different for the images of the two object planes.

3. Since the amount of translation depends on the object distance, images obtained while varying rotation angle of the lens exhibit parallax as shown in [Figure 3.13](#). We observe that in each case (of different pupil magnifications), the initially overlapping rows of the two images (from the two object planes) diverge with progressive rotation of the lens.
4. From the above observations, we infer that unless the pupil magnification is equal to one, the inter-image homography is a depth (object distance) dependent perspective mapping. In other words, for every object plane, the inter-image homography is of the form:

$$H(z_o) = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \quad (3.29)$$

5. However, if the pupil magnification is equal to one, then the inter-image homography reduces to a depth dependent scaling transformation of the form:

$$H(z) = \begin{bmatrix} a & 0 & c \\ 0 & a & f \\ 0 & 0 & 1 \end{bmatrix} \quad (3.30)$$

### **3.7.3 Properties of image field induced by lens rotation about the center of the entrance pupil ( $\alpha_x, \alpha_y \in \mathbb{R}, \beta_x, \beta_y = 0; d_e = 0$ )**

Finally, the properties of the image obtained when the lens is rotated about the center of the entrance pupil are depicted in [Figure 3.14](#) and [Figure 3.15](#). We observe that:

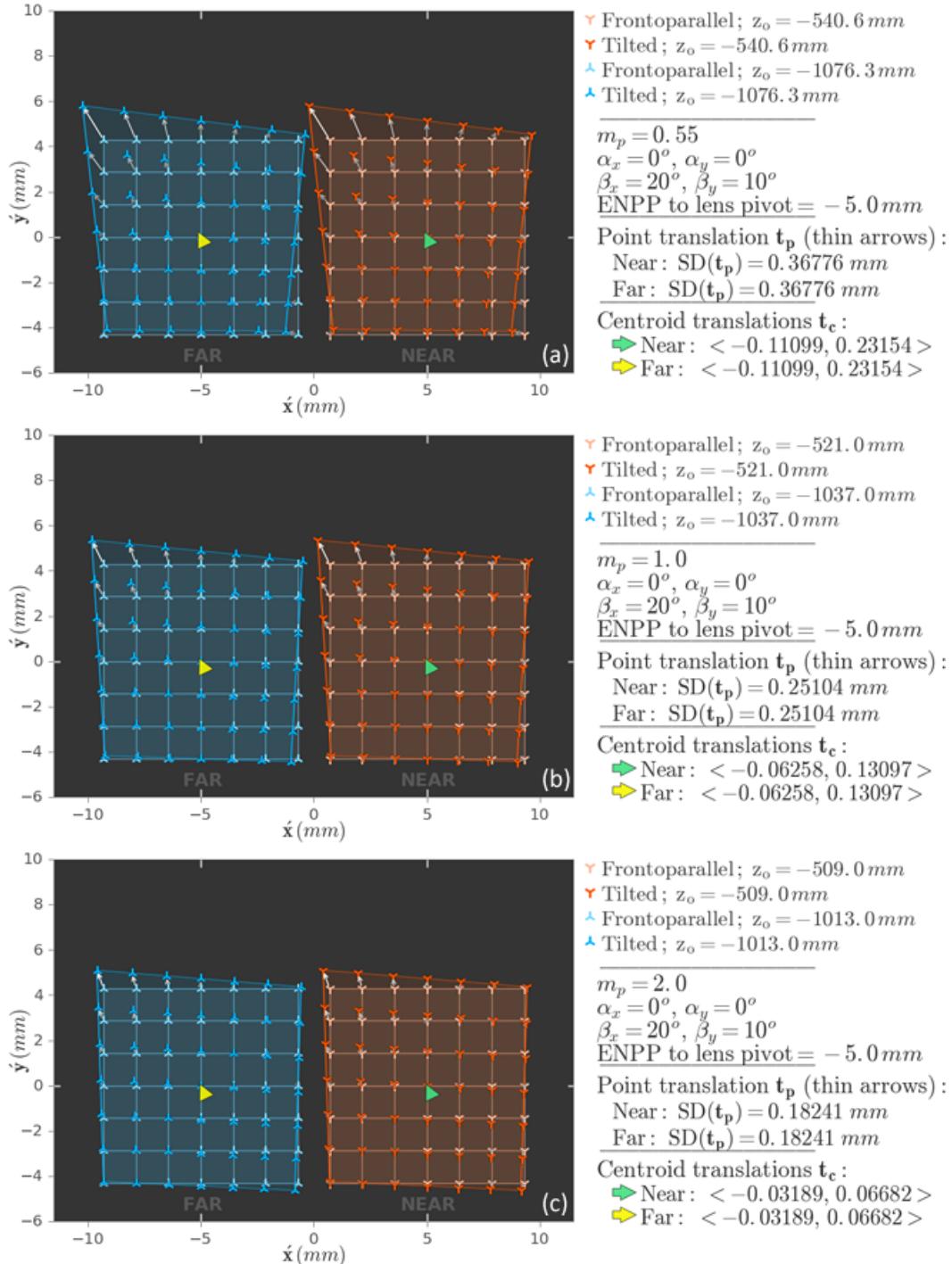
1. Rotation of the lens about the entrance pupil induces a shift of the image field, as must be expected.
2. However, unlike the previous case, the shift of the image field is *independent* of the object distance. We see that the standard deviation of the translation vector lengths is equal for both the group of points (from the two object planes).

3. Since the shift of the image field is independent of the object distance, there is no parallax between images obtained while rotating the lens about the entrance pupil. This is a very important property that can be used for several computational imaging techniques that rely on multiple image capture, including omnifocus (all-in-focus) imaging, digital super-resolution, panoramic imaging, etc.
4. If the pupil magnification is not equal to one, the inter-image homography is a *depth independent* perspective transformation of the form:

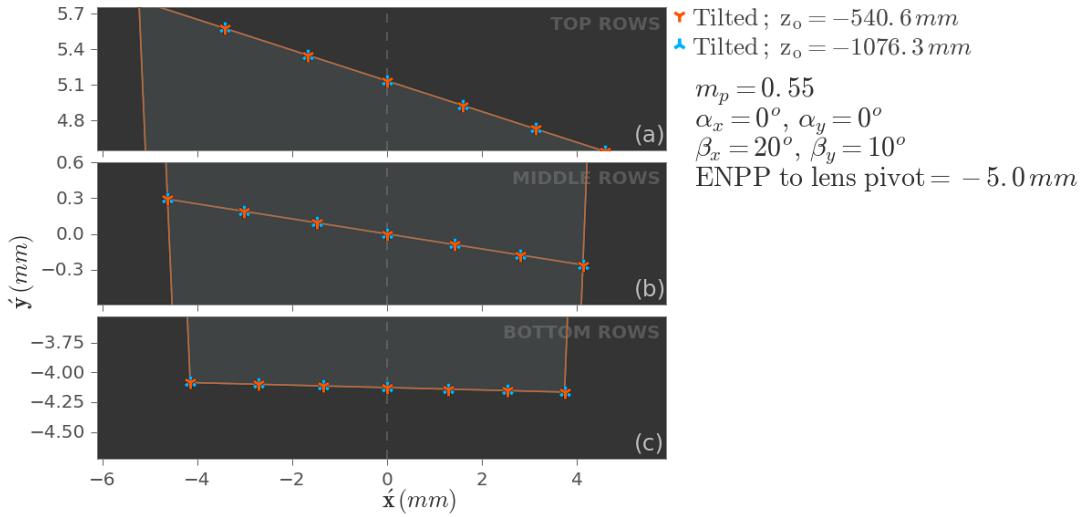
$$H = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \quad (3.31)$$

5. If the pupil magnification is equal to one, then the inter-image homography is a depth independent perspective transformation of the form:

$$H = \begin{bmatrix} a & 0 & c \\ 0 & a & f \\ 0 & 0 & 1 \end{bmatrix} \quad (3.32)$$

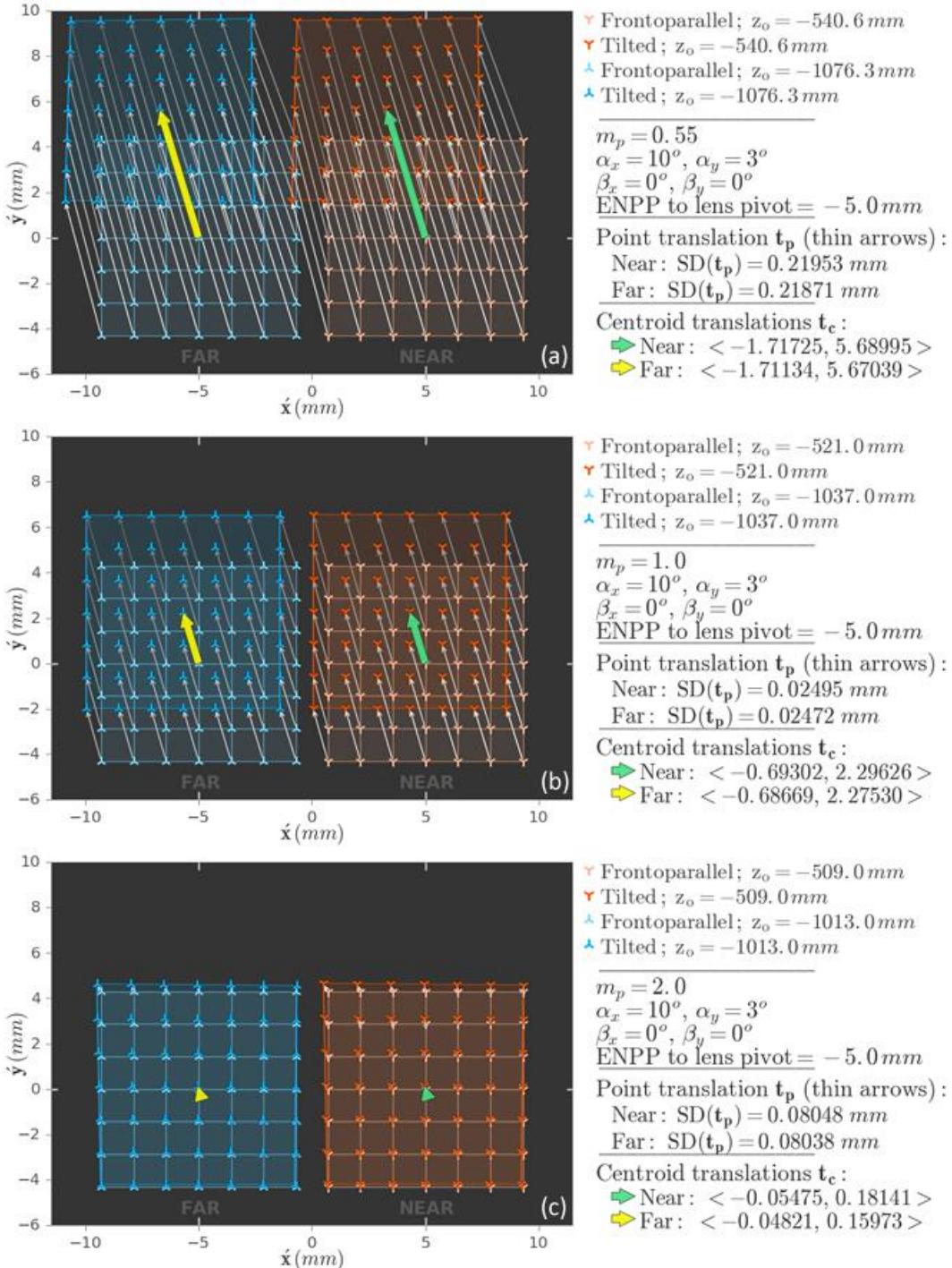


**Figure 3.10** Geometric image under image plane (sensor) rotation for varying pupil magnifications. (a)  $m_p = 0.55$ , (b)  $m_p = 1$ , (c)  $m_p = 2$ .

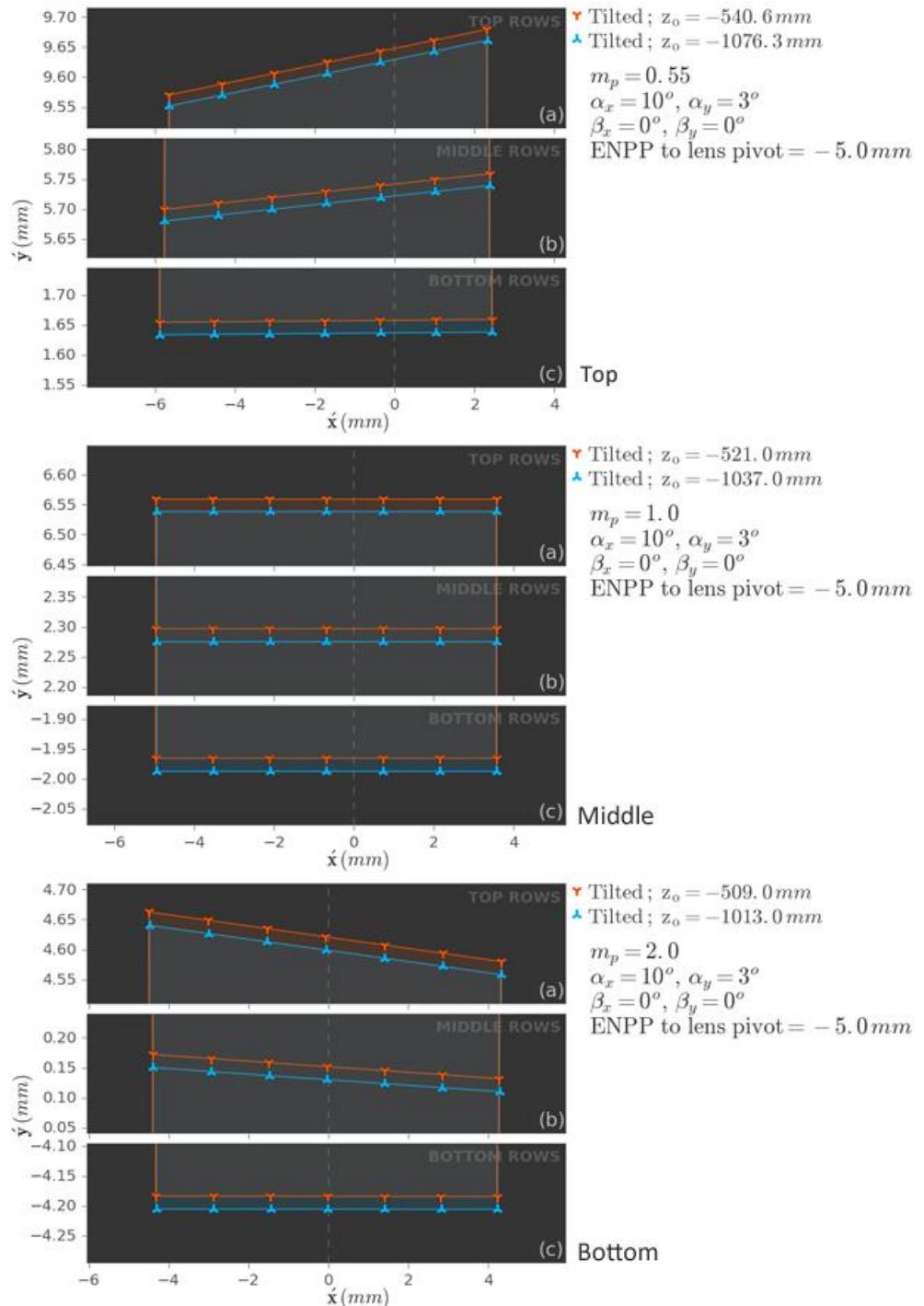


**Figure 3.11** Comparison of geometric distortion induced by sensor rotation for varying object plane distances. The figure shows only the top, middle and bottom rows of image-points for pupil magnification  $m_p = 0.55$  (Figure 3.10(a)). Unlike in the previous figure, the two sets of image-points from the two object planes are left unseparated. We observe that the image-points corresponding to the two object planes that are at different distances from the lens experience the same type and amount of distortion. If that was not the case, then the image-point rows corresponding to the two planes would have diverged.

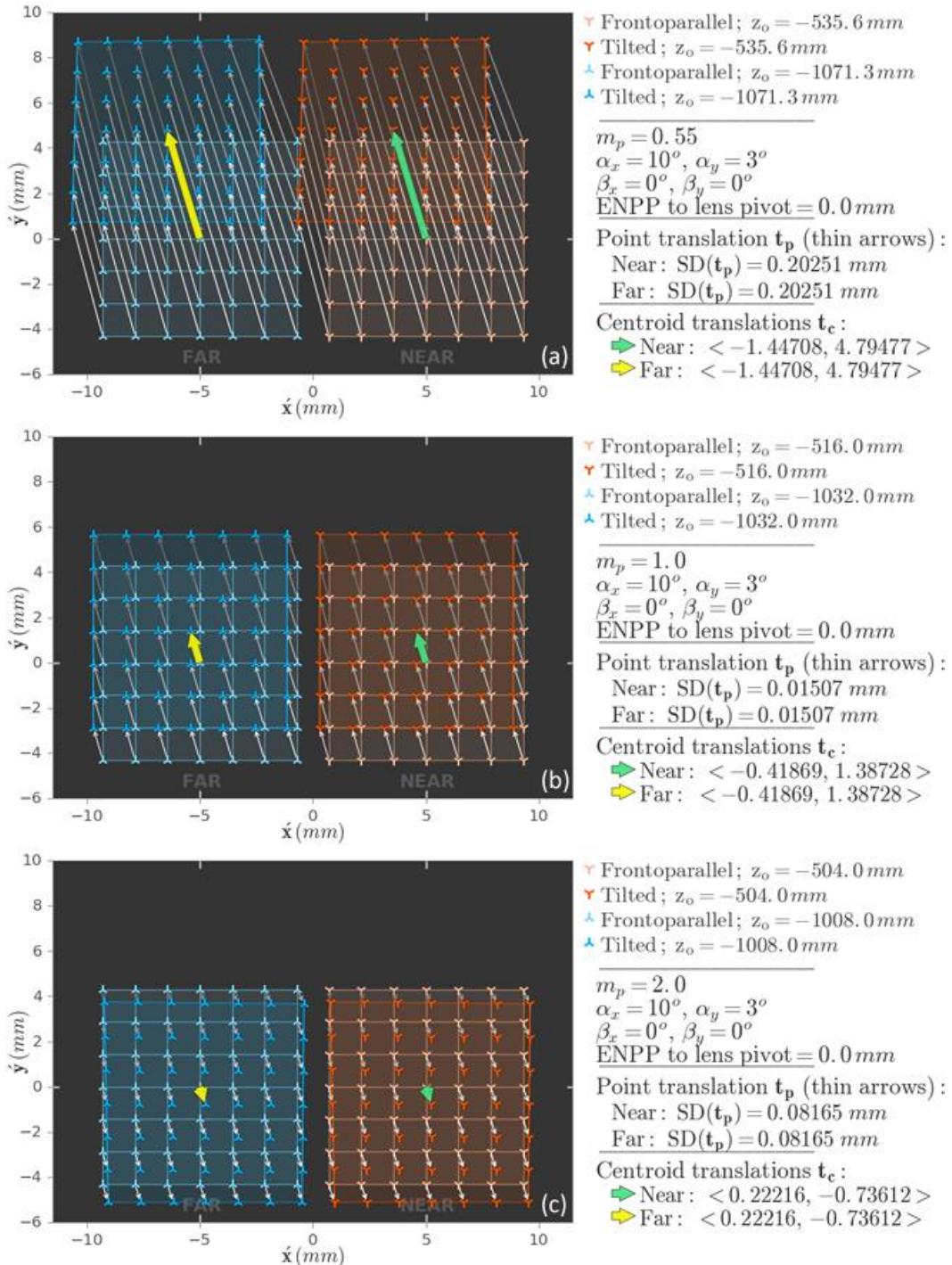
---



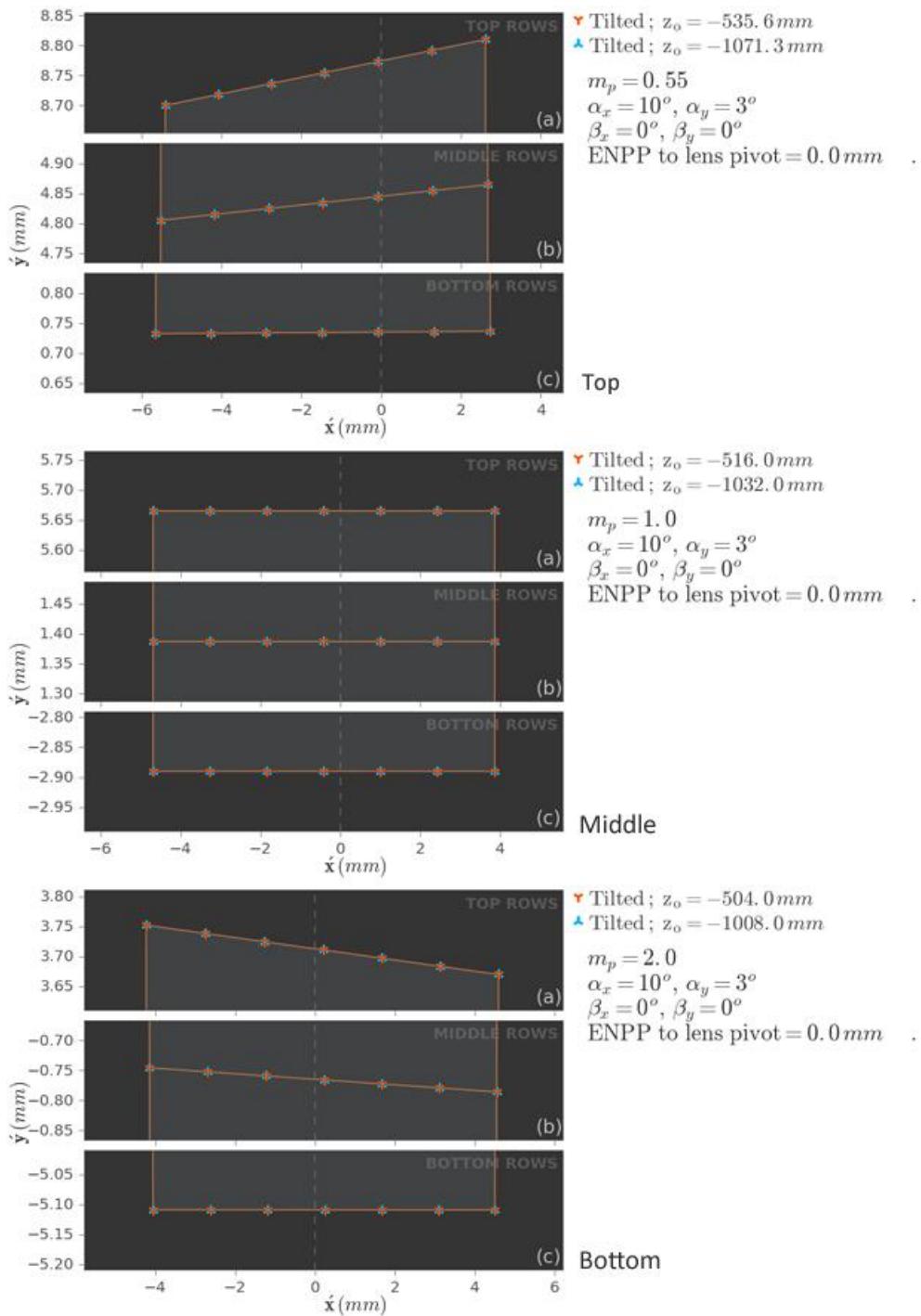
**Figure 3.12** Geometric image under lens rotation away from the entrance pupil for varying pupil magnifications. (a)  $m_p = 0.55$ , (b)  $m_p = 1$ , (c)  $m_p = 2$ .



**Figure 3.13** Variation of geometric distortion of image field induced by lens rotation away from the entrance pupil as a function of object distance and pupil magnification.



**Figure 3.14** Geometric image under lens rotation away from the entrance pupil for varying pupil magnifications. (a)  $m_p = 0.55$ , (b)  $m_p = 1$ , (c)  $m_p = 2$ .



**Figure 3.15** Variation of geometric distortion of images induced by lens rotation *about* the entrance pupil as a function of object distance and pupil magnification.

### 3.8 Summary

The geometric relation between a three-dimensional object point ( ${}^C\mathbf{x}$ ) in camera coordinates and the corresponding image point ( ${}^I\mathbf{x}$ ) in the two-dimensional image plane for an aberration-free Scheimpflug camera is given by Eq. (3.27):

$${}^I\mathbf{x} = R_i^T(\hat{d}_e \mathbf{r}_{\ell,3} - \mathbf{t}_i) + \frac{(\hat{\mathbf{n}}_i(3)\dot{z}_o - \hat{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T ({}^C\mathbf{x} - d_e \mathbf{r}_{\ell,3})} R_i^T R_\ell M_p R_\ell^T ({}^C\mathbf{x} - d_e \mathbf{r}_{\ell,3})$$

This imaging model has two important characteristics that differentiate it from other existing ones:

1. It is a most general geometric imaging relationship between object and image points for rectilinear imaging that accommodates the rotation of the image and the lens planes about their individual pivots. Common imaging models, such as for frontoparallel imaging, directly falls out of the above model.
2. The model incorporates optical parameters such as the pupil distances, pupil locations and pupil magnification, which enable us to accurately analyze the nature of the geometric image obtained the various possible orientation of the lens and image planes.

Our imaging model allowed us to study the effects of rotation of the lens and image planes on the geometric properties of the image. We discovered that if the pupil magnification  $m_p$  equal to one and the lens is rotated about the entrance pupil the inter-image homography—the transformation that relates the image of a scene obtained under varying rotations of the lens—is a simple composition of a scaling and translation transformations.

## Chapter 4

### MODEL OF SCHEIMPFLUG IMAGING – II: FOCUSING

*The scientist explains the world by successive approximations.*

—Erwin Hubble

In the last chapter, we derived a relationship (Eq. (3.27)), between object and image points that allowed us to study and discover important geometric properties of the image in a Scheimpflug camera. However, in deriving Eq. (3.27) we did not impose any constraint on the orientations of the sensor and lens planes that would produce a sharp, focused image of a (tilted) object plane on the sensor plane. Hence, while Eq. (3.27) is useful in analyzing and predicting the geometric properties of the images for a given sensor and lens plane orientation, it is suitable for determining the required orientation of the sensor and lens planes for focusing on a given tilted object plane. In this chapter, we derive a general relationship between the lens, image and object planes that ensures geometric focus on the tilted object plane. Like Eq. (3.27), we aim to explicitly include the pupil parameters in our focusing equation so that we can model the effects of the pupils on focusing. Further, we show (as examples) that this general focusing expression yields object-image relationships that are specific to common types of Scheimpflug configurations.

To keep the model tractable, we impose the constraint that three pivots (that hinge the object, lens, and sensor planes) lie along the  $z$ -axis of the camera frame  $\{C\}$ , and the origin of  $\{C\}$  is co-located with optical axis' pivot (the lens plane is perpendicular to the optical axis). We also

restrict the rotation angles of the object, lens, and image planes between  $-\frac{\pi}{2}$  and  $\frac{\pi}{2}$  about both  $x$ - and  $y$ -axes (in-plane rotations or rotations about the  $z$ -axis is irrelevant for our purpose). Provided we make no distinction between the faces (front or back) of the planes, this restriction on the angles of rotations is not limiting in any way since we can uniquely describe all possible plane orientations in three dimensions. On the other hand, this constraint warrants non-negative values for the  $z$ -component of the plane normals and permits us to estimate the plane normal unambiguously.

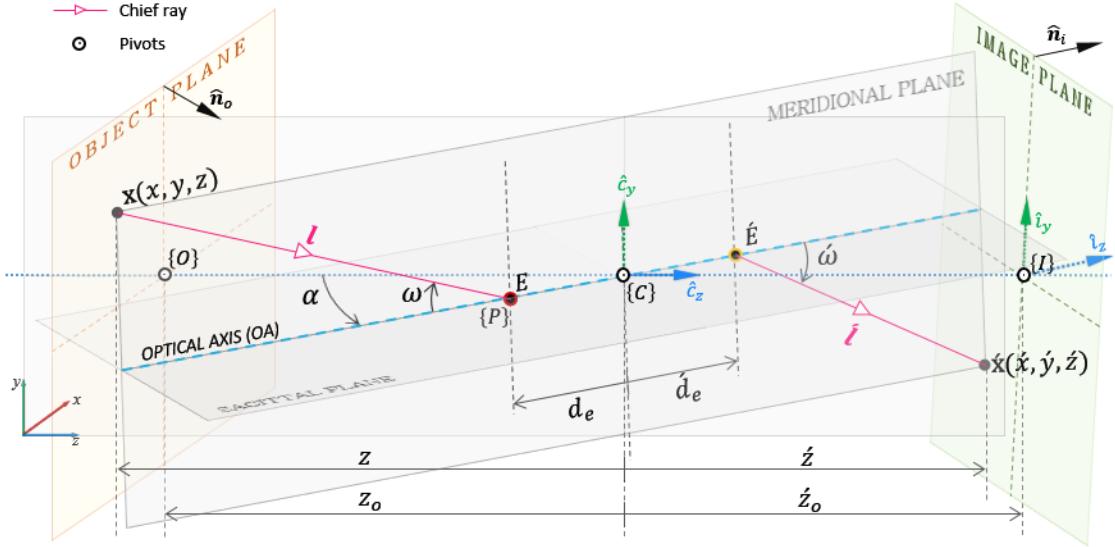
#### 4.1 Relationship between the object, lens, and image planes for focusing

We begin by deriving an expression for the chief ray joining an arbitrary point  $\mathbf{x}$  in the object plane to a point  $\mathbf{x}'$  in the image plane. For  $\mathbf{x}'$  to be the geometric image of  $\mathbf{x}$ , the chief ray between the conjugate points must satisfy the Gaussian imaging equation. This constraint allows us to uniquely determine the position (of the image plane along the  $z$ -axis of  $\{C\}$ ) and orientation of the three planes in Scheimpflug configuration. The setup is shown in [Figure 4.1](#).

The object plane is located at a distance of  $z_o$  (the numerical value of  $z_o$  is negative in our convention) from the origin of camera frame  $\{C\}$ , along the  $z$ -axis. The object plane, pivoted about the point  $(0,0,z_o)$  in the camera frame, is completely described by the pivot point and the normal,  $\hat{\mathbf{n}}_o$ . We describe the object plane normal itself as the product of the rotation matrix  $R_o \in \mathbb{R}^{3 \times 3}$  and  $\hat{\mathbf{c}}_z (= [0,0,1]^T)$ :

$$\hat{\mathbf{n}}_o = R_o \hat{\mathbf{c}}_z. \quad (4.1)$$

The rotation matrix  $R_o$  is typically composed of elementary rotation matrices that represent rotations about the  $x$ - and  $y$ -axis.



**Figure 4.1** Schematic of Scheimpflug imaging. The figure shows the object plane, optical axis, and image plane pivoted about points along the  $z$ -axis,  $\hat{c}_z$ , of the camera frame  $\{C\}$ . The local object plane and image plane coordinates frames ( $\{O\}$  and  $\{I\}$ ) are centered on the object- and image- plane pivots.

If  $d_{\perp}$  be the perpendicular distance of the object plane from the origin, we can determine  $d_{\perp}$  using the Hessian normal form as:

$$\begin{aligned} \hat{n}_o^T \begin{bmatrix} 0 \\ 0 \\ z_o \end{bmatrix} &= d_{\perp}, \\ \text{or } d_{\perp} &= z_o \hat{n}_o^T \hat{c}_z, \end{aligned} \quad (4.2)$$

since  $(0, 0, z_o)$  is a point on the object plane. Further, any general point  $\xi$  on the object plane, satisfies the plane equation:

$$\hat{n}_o^T \xi = z_o (\hat{n}_o^T \hat{c}_z). \quad (4.3)$$

Like the object plane normal, we describe the image plane normal  $\hat{n}_i$  as:

$$\hat{n}_i = R_i \hat{c}_z, \quad (4.4)$$

where  $R_i \in \mathbb{R}^{3 \times 3}$  is a rotation matrix applied to the image plane at the pivot point  $(0, 0, \dot{z}_o)$ .

Repeating the steps used to derive the object plane equation, we obtain the equation for the image plane as:

$$\hat{\mathbf{n}}_i^T \boldsymbol{\xi} = z_o (\hat{\mathbf{n}}_i^T \hat{\mathbf{c}}_z), \quad (4.5)$$

where  $\boldsymbol{\xi}$  is any point on the image plane.

Suppose the entrance ( $E$ ) and exit ( $\acute{E}$ ) pupils are located at distances  $d_e$  and  $\acute{d}_e$  from the lens' pivot (origin of  $\{C\}$ ) respectively, along the optical axis. Also, we describe the rotation of the optical axis by applying the matrix  $R_\ell \in \mathbb{R}^{3 \times 3}$ . Then, the positions of the pupils in  $\{C\}$  following the application of the rotation matrix are:

$$\text{Entrance pupil position, } \mathbf{x}_e = R_\ell [0, 0, d_e]^T = d_e \mathbf{r}_{\ell,3}, \quad (4.6)$$

$$\text{Exit pupil position, } \dot{\mathbf{x}}_e = R_\ell [0, 0, \acute{d}_e]^T = \acute{d}_e \mathbf{r}_{\ell,3}. \quad (4.7)$$

Consider the chief ray from the object point  $\mathbf{x}$  to the corresponding image point  $\dot{\mathbf{x}}$ , passing through the center of the entrance and exit pupils. Let the direction cosines of the ray in the object- and image-space be  $\mathbf{l}$  and  $\dot{\mathbf{l}}$  respectively. Since  $\mathbf{l}$  is the (unit-length) direction vector of the ray from  $\mathbf{x}$  to  $\mathbf{x}_e$ , we can write:

$$\mathbf{x}_e = \mathbf{x} + \underline{y} \mathbf{l}, \quad (4.8)$$

where,  $\mathbf{x}_e$  is the position of the entrance pupil center, and  $\underline{y}$  is the length of the ray. Substituting Eq. (4.6) into Eq. (4.8), we obtain:

$$\mathbf{x} = d_e \mathbf{r}_{\ell,3} - \underline{y} \mathbf{l}. \quad (4.9)$$

Further, since  $\mathbf{x}$  is a point on the object plane, it satisfies the object plane Eq. (4.3). Substituting  $\mathbf{x}$  into Eq. (4.3) and rearranging terms we obtain:

$$u = \frac{-z_o(\hat{\mathbf{n}}_o^T \hat{\mathbf{c}}_z) + d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_o^T \mathbf{l}} . \quad (4.10)$$

The chief ray in the image space emerges from the exit pupil with a direction cosine vector  $\mathbf{l}$ . Therefore, we can write the equation of the chief ray in parametric form (using Eq. (4.7)) as:

$$\xi = \mathbf{x}_e + \lambda(\mathbf{l}) = \hat{d}_e \mathbf{r}_{\ell,3} + \lambda \mathbf{l} , \quad (4.11)$$

where  $\xi$  is any point on the ray,  $\mathbf{x}_e$  is the position of the exit pupil, and  $\lambda$  is the length of the ray.

If the length of the chief ray in the image space be  $\dot{u}$ , then at the point of intersection of the chief ray with the image plane  $\xi = \mathbf{x}$  and  $\lambda = \dot{u}$  in Eq. (4.11). Therefore, we obtain

$$\mathbf{x} = \hat{d}_e \mathbf{r}_{\ell,3} + \dot{u} \mathbf{l} . \quad (4.12)$$

Substituting  $\mathbf{x}$  in Eq. (4.5) and rearranging terms, we obtain

$$\dot{u} = \frac{z_o(\hat{\mathbf{n}}_i^T \hat{\mathbf{c}}_z) - \hat{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T \mathbf{l}} . \quad (4.13)$$

Eqs. (4.10) and (4.13) gives the length of the chief rays between  $\mathbf{x}$  and  $\mathbf{x}_e$  in the object space and between  $\mathbf{x}_e$  and  $\mathbf{x}$  in the image space respectively. For  $\mathbf{x}$  to be the geometrically focused image of  $\mathbf{x}$ , the lengths in the object and image space must satisfy the Gaussian imaging equation.

The well-known Gaussian imaging equation ( $-1/u + 1/\dot{u} = 1/f$ ) relates the focal length and the conjugate plane *directed* distances measured from the principal planes. Instead, if the directed distances are specified with respect to the pupil planes (from entrance pupil to object plane; and from exit pupil to image plane), then a variant of the Gaussian imaging equation is used, which incorporates the pupil magnification  $m_p$  into the formula:

$$-\frac{1}{m_p u} + \frac{m_p}{\dot{u}} = \frac{1}{f} , \quad (4.14)$$

where  $u$  and  $\dot{u}$  are directed distances along the optical axis measured from the entrance pupil to the object plane, and from the exit pupil to the image plane respectively;  $m_p$  is the pupil magnification; and  $f$  is the focal length. We have provided a derivation and a brief exposition of Eq. (4.14) in [Appendix B.1](#).

The most common application of Eq. (4.14) is for frontoparallel imaging in which the conjugate planes are parallel to each other and perpendicular to the optical axis. Moreover, *all* pairs of object-image conjugate points satisfy this relation even if the ensemble of object- and image-points belong to planes on object and image sides respectively that are *not parallel* to each other.

The ray vector of length  $\underline{u}$  and direction  $\mathbf{l}$  in the object space is  $\underline{u}\mathbf{l}$ . The projection of this ray vector on the optical axis  $\hat{\mathbf{o}} (= R_\ell \hat{\mathbf{c}}_z)$  is  $u = \underline{u}(\mathbf{l} \cdot \hat{\mathbf{o}})$ . Similarly, the ray projection of the image space ray vector on the optical axis is  $\dot{u} = \dot{\underline{u}}(\hat{\mathbf{l}} \cdot \hat{\mathbf{o}})$ . In order to substitute  $u$  and  $\dot{u}$  in to Eq. (4.14) we need to ensure that they are the directed distances. Following our sign convention, the directed distance from the entrance pupil to the object point is  $-u$ . Substituting  $-u$  and  $\dot{u}$  into Eq. (4.14) we obtain:

$$\frac{1}{m_p \underline{u}(\mathbf{l} \cdot \hat{\mathbf{o}})} + \frac{m_p}{\dot{\underline{u}}(\hat{\mathbf{l}} \cdot \hat{\mathbf{o}})} = \frac{1}{f}. \quad (4.15)$$

Further, substituting the expressions for  $\underline{u}$  and  $\dot{\underline{u}}$  (Eqs. (4.10) and (4.13)) into the above equation, we obtain:

$$-\frac{\hat{\mathbf{n}}_o^T \mathbf{l}}{m_p [z_o(\hat{\mathbf{n}}_o^T \hat{\mathbf{c}}_z) - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})](\mathbf{l} \cdot \hat{\mathbf{o}})} + \frac{m_p \hat{\mathbf{n}}_i^T \hat{\mathbf{l}}}{[\dot{z}_o(\hat{\mathbf{n}}_i^T \hat{\mathbf{c}}_z) - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})](\hat{\mathbf{l}} \cdot \hat{\mathbf{o}})} = \frac{1}{f}. \quad (4.16)$$

The direction cosine of the chief ray in the image space,  $\hat{\mathbf{l}}$ , is related to the direction cosine of the chief ray in the object space as (Eq. (3.12)):

$$\hat{\boldsymbol{l}} = \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} R_\ell M_p R_\ell^T \boldsymbol{l} .$$

Substituting  $\hat{\boldsymbol{l}}$ , into Eq. (4.16) we obtain:

$$\begin{aligned} & - \frac{\hat{\boldsymbol{n}}_o^T \boldsymbol{l}}{m_p [z_o(\hat{\boldsymbol{n}}_o^T \hat{\boldsymbol{c}}_z) - d_e(\hat{\boldsymbol{n}}_o^T \boldsymbol{r}_{\ell,3})] (\boldsymbol{l} \cdot \hat{\boldsymbol{o}})} \\ & + \frac{m_p \hat{\boldsymbol{n}}_i^T (R_\ell M_p R_\ell^T \boldsymbol{l})}{[\dot{z}_o(\hat{\boldsymbol{n}}_i^T \hat{\boldsymbol{c}}_z) - \dot{d}_e(\hat{\boldsymbol{n}}_i^T \boldsymbol{r}_{\ell,3})] (R_\ell M_p R_\ell^T \boldsymbol{l}) \cdot \hat{\boldsymbol{o}}} = \frac{1}{f} . \end{aligned} \quad (4.17)$$

To simplify the above expression, let us consider  $(R_\ell M_p R_\ell^T \hat{\boldsymbol{n}}_i)^T \boldsymbol{l}$  as

$$\begin{aligned} (R_\ell M_p R_\ell^T \hat{\boldsymbol{n}}_i)^T \boldsymbol{l} &= \hat{\boldsymbol{n}}_i^T (R_\ell^T)^T (M_p)^T R_\ell^T \boldsymbol{l} \\ &= \hat{\boldsymbol{n}}_i^T R_\ell M_p R_\ell^T \boldsymbol{l} . \quad (\because M_p \text{ is a diagonal matrix} \\ & \quad \text{and } R_\ell \text{ is a rotation matrix}) \end{aligned} \quad (4.18)$$

Therefore, we can write  $\hat{\boldsymbol{n}}_i^T (R_\ell M_p R_\ell^T \boldsymbol{l})$  in Eq. (4.17) as  $(R_\ell M_p R_\ell^T \hat{\boldsymbol{n}}_i)^T \boldsymbol{l}$ . Similarly, we can also write  $(R_\ell M_p R_\ell^T \boldsymbol{l}) \cdot \hat{\boldsymbol{o}}$  as  $(R_\ell M_p R_\ell^T \boldsymbol{l})^T \hat{\boldsymbol{o}} = \hat{\boldsymbol{o}}^T (R_\ell M_p R_\ell^T \boldsymbol{l}) = (R_\ell M_p R_\ell^T \hat{\boldsymbol{o}})^T \boldsymbol{l}$ . Then, Eq. (4.17) can be written as:

$$\begin{aligned} & - \frac{\hat{\boldsymbol{n}}_o^T \boldsymbol{l}}{m_p [z_o(\hat{\boldsymbol{n}}_o^T \hat{\boldsymbol{c}}_z) - d_e(\hat{\boldsymbol{n}}_o^T \boldsymbol{r}_{\ell,3})] \hat{\boldsymbol{o}}^T \boldsymbol{l}} + \frac{m_p (R_\ell M_p R_\ell^T \hat{\boldsymbol{n}}_i)^T \boldsymbol{l}}{[\dot{z}_o(\hat{\boldsymbol{n}}_i^T \hat{\boldsymbol{c}}_z) - \dot{d}_e(\hat{\boldsymbol{n}}_i^T \boldsymbol{r}_{\ell,3})] (R_\ell M_p R_\ell^T \hat{\boldsymbol{o}})^T \boldsymbol{l}} \\ & = \frac{1}{f} . \end{aligned} \quad (4.19)$$

We can further simplify the above equation by noting that:

1.  $\hat{\boldsymbol{n}}_o^T \hat{\boldsymbol{c}}_z = \hat{\boldsymbol{n}}_o^T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \hat{\boldsymbol{n}}_o(3) ,$
2.  $\hat{\boldsymbol{o}}^T \boldsymbol{l} = (R_\ell \hat{\boldsymbol{c}}_z)^T \boldsymbol{l} = \boldsymbol{r}_{\ell,3}^T \boldsymbol{l} ,$

3.  $\hat{\mathbf{n}}_i^T \hat{\mathbf{c}}_z = \hat{\mathbf{n}}_i^T \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \hat{\mathbf{n}}_i(3)$ , and
4.  $(R_\ell M_p R_\ell^T \hat{\mathbf{o}})^T \mathbf{l} = (R_\ell M_p R_\ell^T R_\ell \hat{\mathbf{c}}_z)^T \mathbf{l} = (R_\ell M_p \hat{\mathbf{c}}_z)^T \mathbf{l} = m_p \mathbf{r}_{\ell,3}^T \mathbf{l}$ .

Using the above results Eq. (4.19) reduces to:

$$-\frac{\hat{\mathbf{n}}_o^T \mathbf{l}}{m_p [z_o \hat{\mathbf{n}}_o(3) - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})] \mathbf{r}_{\ell,3}^T \mathbf{l}} + \frac{(R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i)^T \mathbf{l}}{[\dot{z}_o \hat{\mathbf{n}}_i(3) - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})] \mathbf{r}_{\ell,3}^T \mathbf{l}} = \frac{1}{f}. \quad (4.20)$$

Further, multiplying the above equation by the scalar  $\mathbf{r}_{\ell,3}^T \mathbf{l}$ , and using the commutative and distributive properties of dot product, we obtain:

$$\mathbf{l}^T \left[ -\frac{\hat{\mathbf{n}}_o}{m_p [z_o \hat{\mathbf{n}}_o(3) - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})]} + \frac{R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i}{[\dot{z}_o \hat{\mathbf{n}}_i(3) - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})]} - \frac{\mathbf{r}_{\ell,3}}{f} \right] = 0. \quad (4.21)$$

Note that in the above expressions  $\mathbf{r}_{\ell,3}$ , the third column of the rotation matrix  $R_\ell$ , is the unit vector along the optical axis.

The direction cosine vector  $\mathbf{l}$  has  $\ell^2$ -Norm equal to one. Therefore, the above equation is satisfied only if the second vector,  $\left[ -\frac{\hat{\mathbf{n}}_o}{m_p [z_o \hat{\mathbf{n}}_o(3) - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})]} + \frac{R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i}{[\dot{z}_o \hat{\mathbf{n}}_i(3) - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})]} - \frac{\mathbf{r}_{\ell,3}}{f} \right]$ , is either perpendicular to  $\mathbf{l}$  or identically equal to zero. Now, is it possible for the second vector to be perpendicular to  $\mathbf{l}$  (the chief ray's direction vector)? As we did not make any specific assumptions about  $\mathbf{l}$  in the derivation of Eq. (4.21), *all* chief rays—an infinitude of  $\mathbf{l}$  vectors within the object-and image-space perspective cones—must satisfy Eq. (4.21). Are all possible  $\mathbf{l}$  vectors perpendicular to the second vector? Since the second vector is a linear combination of  $\hat{\mathbf{n}}_o$  (the object plane normal),  $R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i$  (the transformed image plane normal), and  $\mathbf{r}_{\ell,3}$  (unit vector along the optical axis), we can deduce that  $\mathbf{l}$ , *in general*, is not perpendicular to the second vector. Therefore, the second vector must be equal to zero, which yields:

$$-\frac{\hat{\mathbf{n}}_o}{m_p[z_o\hat{\mathbf{n}}_o(3) - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})]} + \frac{R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i}{[\dot{z}_o \hat{\mathbf{n}}_i(3) - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})]} = \frac{\mathbf{r}_{\ell,3}}{f}. \quad (4.22)$$

Further, we can simplify Eq. (4.22) if we let  $\hat{\mathbf{n}}_o = \frac{\hat{\mathbf{n}}_o}{\hat{\mathbf{n}}_o(3)} = \left[ \frac{\hat{\mathbf{n}}_o(1)}{\hat{\mathbf{n}}_o(3)}, \frac{\hat{\mathbf{n}}_o(2)}{\hat{\mathbf{n}}_o(3)}, 1 \right]^T$  and  $\hat{\mathbf{n}}_i = \frac{\hat{\mathbf{n}}_i}{\hat{\mathbf{n}}_i(3)} = \left[ \frac{\hat{\mathbf{n}}_i(1)}{\hat{\mathbf{n}}_i(3)}, \frac{\hat{\mathbf{n}}_i(2)}{\hat{\mathbf{n}}_i(3)}, 1 \right]^T$ . Then, after factoring  $\hat{\mathbf{n}}_o(3)$  and  $\hat{\mathbf{n}}_i(3)$  out of the denominator terms, we can write Eq. (4.22) as

$$\boxed{-\frac{\hat{\mathbf{n}}_o}{m_p[z_o - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})]} + \frac{R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i}{[\dot{z}_o - \dot{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})]} = \frac{\mathbf{r}_{\ell,3}}{f}}. \quad (4.23)$$

Where,

$z_o$	Directed distance of the object plane's pivot from the origin of the camera frame, along the z-axis of the camera frame. (In most of the problems that we are interested, the value of $z_o$ is negative.)
$\dot{z}_o$	Directed distance of the image plane's pivot from the camera frame, along the z-axis of the camera frame.
$f$	Focal length of the lens.
$M_p$	Equal to $\text{diag}(1, 1, m_p)$ , where $m_p$ is the pupil magnification.
$d_e$	Location of the entrance pupil from the pivot (origin of $\{C\}$ ) along the optical axis. This scalar quantity physical units, usually in millimeters.
$\dot{d}_e$	Location of the exit pupil from the pivot (origin of $\{C\}$ ) along the optical axis.
$R_\ell$	Rotation matrix used to describe the orientation of the lens plane.
$\mathbf{r}_{\ell,3}$	The third column of $R_\ell$ .
$\hat{\mathbf{n}}_o$	Equals $\hat{\mathbf{n}}_o / \hat{\mathbf{n}}_o(3)$ , where $\hat{\mathbf{n}}_o$ is the object plane normal.
$\hat{\mathbf{n}}_i$	Equals $\hat{\mathbf{n}}_i / \hat{\mathbf{n}}_i(3)$ , where $\hat{\mathbf{n}}_i$ is the image plane normal.

The expedient simplification from Eq. (4.22) to Eq. (4.23) relies on our ability to describe the *unit* normal vectors  $\hat{\mathbf{n}}_o$  and  $\hat{\mathbf{n}}_i$  using only the components along  $x$ - and  $y$ -axes. If we know the

$x$ - and  $y$ - components of the normal vector, we can determine the  $z$ - component uniquely because we have restricted the angles of plane rotations between  $-\frac{\pi}{2}$  and  $\frac{\pi}{2}$  about both  $x$ - and  $y$ -axes (one of the starting assumptions). For example, if the object- and lens-plane orientations and distances are known, and we estimate the image plane distance  $z_o$  and orientation vector  $\hat{\mathbf{n}}_i (= [\dot{p}, \dot{q}, 1]^T )$  of the image plane using Eq. (4.23), then we can determine the image plane normal as:

$$\begin{aligned}\hat{\mathbf{n}}_i(3) &= \frac{1}{\sqrt{\dot{p}^2 + \dot{q}^2 + 1}} , \\ \hat{\mathbf{n}}_i(1) &= \hat{\mathbf{n}}_i(3) \dot{p} , \\ \hat{\mathbf{n}}_i(2) &= \hat{\mathbf{n}}_i(3) \dot{q} ,\end{aligned}\tag{4.24}$$

where  $\dot{p} = \frac{\hat{\mathbf{n}}_i(1)}{\hat{\mathbf{n}}_i(3)}$ ,  $\dot{q} = \frac{\hat{\mathbf{n}}_i(2)}{\hat{\mathbf{n}}_i(3)}$ , and we have dropped the negative sign from the expression for  $\hat{\mathbf{n}}_i(3)$  since  $\hat{\mathbf{n}}_i(3)$  is guaranteed to be positive as discussed under the assumptions at the beginning of this section.

Eq. (4.23) is most general in the sense that it readily yields the various expressions for specific cases of object, lens and image plane orientations as we now demonstrate in the following set of examples.

## 4.2 Examples of typical Scheimpflug imaging configurations

### 4.2.1 Example: Focusing in frontoparallel configuration

Suppose the object plane is frontoparallel with the lens plane associated with a thick lens and  $z_o$  units from the origin of camera frame  $\{C\}$  (along the  $z$ -axis of the  $\{C\}$ ), what should be the orientation and position of the image plane to focus on the object plane?

Based on the given data, we have the following:

1. Since the lens is not tilted,  $R_\ell = \mathbf{I}$ ,  $\mathbf{r}_{\ell,3} = [0, 0, 1]^T$ .
2. Since the object plane is not tilted,  $\hat{\mathbf{n}}_o = [0, 0, 1]^T$  and  $\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3} = 1$ .

3. Let  $\dot{p} = \frac{\hat{n}_i(1)}{\hat{n}_i(3)}$  and  $q = \frac{\hat{n}_i(2)}{\hat{n}_i(3)}$ . Then,  $\hat{\mathbf{n}}_i = [\dot{p}, \dot{q}, 1]^T$ , and  $\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3} = 1$ , and  $R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i = [\dot{p}, \dot{q}, m_p]^T$

Substituting the above parameters in Eq. (4.23) we have

$$-\frac{1}{m_p[z_o - d_e]} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{[\dot{z}_o - \dot{d}_e]} \begin{bmatrix} \dot{p} \\ \dot{q} \\ m_p \end{bmatrix} = \frac{1}{f} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad (4.25)$$

which yields the following relations from each row of Eq. (4.25):

$$\dot{p} = 0, \quad (4.26)$$

$$\dot{q} = 0, \quad (4.27)$$

$$\dot{z}_o = \dot{d}_e + \frac{m_p^2 f (z_o - d_e)}{m_p (z_o - d_e) + f}. \quad (4.28)$$

Writing  $z_e = (z_o - d_e)$  and  $\dot{z}_e = (\dot{z}_o - \dot{d}_e)$  where,  $z_e$  is the directed distance from the entrance pupil to the object plane and  $\dot{z}_e$  is the directed distance from the exit pupil to the image plane, we can rewrite Eq. (4.28) as

$$\dot{z}_e = \frac{m_p^2 f z_e}{m_p z_e + f}. \quad (4.29)$$

As a concrete example, if  $m_p = 2$ ,  $f = 24 \text{ mm}$ ,  $d_e = -5 \text{ mm}$ ,  $\dot{d}_e = -25 \text{ mm}$ , and  $z_o = -509.0 \text{ mm}$ , then we can calculate the image plane distance from the camera frame's origin using Eq. (4.28) to be  $24.1707 \text{ mm}$ .

Further, if  $|m_p z_e| > |f|$ , then the sign of  $\dot{z}_e$  is positive, which is a condition for real and inverted images. For example, if  $z_e = 1 \text{ m}$ , and  $f = 100 \text{ mm}$ , then  $\dot{z}_e$  is positive for  $m_p > 0.1$ . If the sign of  $\dot{z}_e$  is negative, then a virtual and upright image is formed in front of the lens. In a survey of 120 imaging lenses (see [Appendix B.2](#)) from a database (Zemax Zebase) of well-designed lenses,

we found over 90% of all lenses to have pupil magnification greater than 0.5 and no lens having pupil magnification less than 0.2. Thus, in the common imaging scenarios, the sign of  $\dot{z}_e$  is positive.

From Eqs. (4.26) and (4.27) we see that the image plane normal  $\hat{\mathbf{n}}_i$  is equal to  $[0, 0, 1]^T$ . This implies that the image plane is parallel to the lens and object plane and perpendicular to the optical axis if the object plane is parallel to the lens plane.

Additionally, if a thin lens model ( $m_p = 1, d_e = \dot{d}_e = 0$ ), Eq. (4.28) reduces to the Gaussian lens equation for thin lenses:

$$\dot{z}_o = \frac{z_o f}{z_o + f}. \quad (4.30)$$

Furthermore, if the distances to the object and image plane are specified from the object- and image- space principal planes instead of the pupils, Eq. (4.28) reduces to ( $m_p = 1$ )

$$\dot{z}_o = \dot{d}_e + \frac{(z_o - d_e)f}{(z_o - d_e) + f}, \quad (4.31)$$

in which,  $(z_o - d_e)$  is the distance of the object plane from the object-space principal point, and  $d_e$  and  $\dot{d}_e$  represent the locations of the object- and image-space principal points in the camera frame  $\{C\}$ . While  $z_o$  is numerically negative, the signs of  $d_e$  and  $\dot{d}_e$  depend on the position of the principal points with respect to the origin of  $\{C\}$ .

#### 4.2.2 Example: Focusing on tilted object plane by tilting the image plane

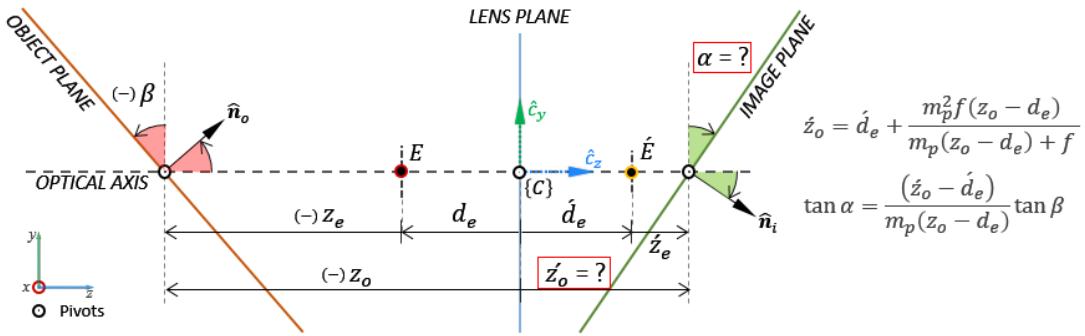
Suppose the object plane, pivoted at  $(0, 0, -z_o)$ , is tilted by an angle  $\beta$  about the  $x$ -axis ([Figure 4.2](#)), and the lens plane perpendicular to the  $z$ -axis of the camera frame  $\{C\}$ , what is the conjugate orientation and position of the image plane for achieving a geometrically focused image assuming a thick lens model?

1. Since the lens is not tilted,  $R_\ell = \mathbf{I}, \mathbf{r}_{\ell,3} = [0, 0, 1]^T$ .
2. Let  $p = \frac{\hat{\mathbf{n}}_o(1)}{\hat{\mathbf{n}}_o(3)}$  and  $q = \frac{\hat{\mathbf{n}}_o(2)}{\hat{\mathbf{n}}_o(3)}$ . Then,  $\hat{\mathbf{n}}_o = [p, q, 1]^T$  and  $\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3} = 1$ .

3. Let  $\dot{p} = \frac{\hat{n}_i(1)}{\hat{n}_i(3)}$  and  $\dot{q} = \frac{\hat{n}_i(2)}{\hat{n}_i(3)}$ . Then,  $\hat{n}_i = [\dot{p}, \dot{q}, 1]^T$ ,  $\hat{n}_i^T \mathbf{r}_{\ell,3} = 1$  and  $R_\ell M_p R_\ell^T \hat{n}_i = [\dot{p}, \dot{q}, m_p]^T$ .

Substituting the above parameters in Eq. (4.23) we obtain:

$$-\frac{1}{m_p[z_o - d_e]} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} + \frac{1}{[\dot{z}_o - \dot{d}_e]} \begin{bmatrix} \dot{p} \\ \dot{q} \\ m_p \end{bmatrix} = \frac{1}{f} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (4.32)$$



**Figure 4.2** Object and image plane tilt. In the above cross-sectional (y-z plane) view, the object plane is tilted by an angle  $\beta$  about the x-axis at  $(0, 0, -z_o)$ . We would like to find the position and orientation of the image plane in order to focus on the object.

---

As in the [Example 4.3.1](#), we obtain the distance of the image plane pivot in the camera frame  $\{C\}$ , from the third row of Eq. (4.32) as:

$$\dot{z}_o = \dot{d}_e + \frac{m_p^2 f(z_o - d_e)}{m_p(z_o - d_e) + f}. \quad (4.33)$$

Also, as in [Example 4.3.1](#), writing  $z_e = (z_o - d_e)$  and  $\dot{z}_e = (\dot{z}_o - \dot{d}_e)$  where,  $z_e$  is the directed distance from the entrance pupil to the origin of the object plane and  $\dot{z}_e$  is the directed distance from the exit pupil to the origin of the image plane, we obtain:

$$\dot{z}_e = \frac{m_p^2 f z_e}{m_p z_e + f} . \quad (4.34)$$

in which  $z_e$  is numerically negative and  $\dot{z}_e$  is positive for macroscopic imaging.

From the first and second rows of Eq. (4.32), we obtain:

$$\dot{p} = \frac{(\dot{z}_o - \dot{d}_e)}{m_p(z_o + d_e)} p , \quad (4.35)$$

and

$$\dot{q} = \frac{(\dot{z}_o - \dot{d}_e)}{m_p(z_o + d_e)} q . \quad (4.36)$$

Since the object plane is rotated by  $\beta$  about the  $x$ -axis, the rotation matrix  $R_o \in \mathbb{R}^{3 \times 3}$  is

$$R_o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix} . \quad (4.37)$$

Note that if the direction of rotation of the object plane about the  $x$ -axis is from  $+z$ -axis to  $+y$ -axis (as depicted in [Figure 4.2](#)), then  $\beta$  is numerically negative.

The object plane normal is (Eq. (4.1)) is:

$$\hat{\mathbf{n}}_o = R_o \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\sin \beta \\ \cos \beta \end{bmatrix} , \quad (4.38)$$

from which, we get  $\hat{\mathbf{n}}_o$  (by dividing  $\hat{\mathbf{n}}_o$  by  $\hat{\mathbf{n}}_o(3)$ ) as:

$$\hat{\mathbf{n}}_o = \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\tan \beta \\ 1 \end{bmatrix} . \quad (4.39)$$

Substituting  $\hat{\mathbf{n}}_o$  into Eqs. (4.35) and (4.36) we obtain:

$$\dot{p} = 0 \quad \text{and} \quad \dot{q} = -\frac{(\dot{z}_o - \dot{d}_e)}{m_p(z_o - d_e)} \tan \beta . \quad (4.40)$$

Eq. (4.40) in conjunction with Eq. (4.24) suggest that  $\hat{\mathbf{n}}_i = \langle 0, \frac{\dot{q}}{\sqrt{1+\dot{q}}}, \frac{1}{\sqrt{1+\dot{q}}} \rangle$ , which implies that the image plane normal is confined to the  $y$ - $z$  plane. That is, if the object plane is rotated only about the  $x$ -axis, then the image plane must also be rotated only about the  $x$ -axis to achieve geometric focus on the tilted object plane. The exact expression for the components of the image plane normal is obtained by substituting  $\dot{p}$  and  $\dot{q}$  into Eq. (4.24). The angle of the image plane normal with respect to the three axes is further obtained as the cosine inverses of the corresponding components of the normal vector.

Furthermore, there remains a desideratum to have a more direct relationship between the rotation angles of the object and image planes that can be readily used instead of computing the plane normals. To that end, let us suppose that the required image plane tilt (rotation about the  $x$ -axis) is  $\alpha$ . If we represent the rotation matrix of the image plane as:

$$R_i = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} , \quad (4.41)$$

we obtain the image plane normal  $\hat{\mathbf{n}}_i$  as:

$$\hat{\mathbf{n}}_i = R_i \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\sin \alpha \\ \cos \alpha \end{bmatrix} , \quad (4.42)$$

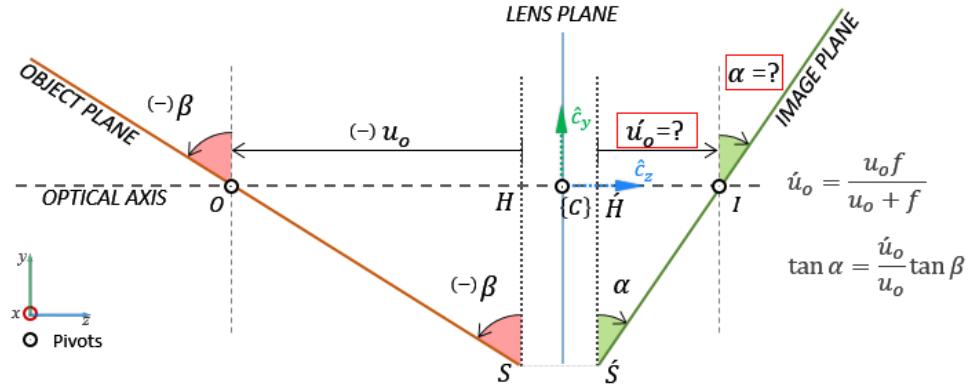
which produces  $\hat{\mathbf{n}}_i$  as:

$$\hat{\mathbf{n}}_i = \begin{bmatrix} \dot{p} \\ \dot{q} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\tan \alpha \\ 1 \end{bmatrix} . \quad (4.43)$$

Substituting  $\dot{q} = -\tan \alpha$  (from Eq. (4.43)) into Eq. (4.40) we obtain:

$$\tan \alpha = \frac{(\dot{z}_o - \dot{d}_e)}{m_p(z_o - d_e)} \tan \beta = \frac{\dot{z}_e}{m_p z_e} \tan \beta \quad (4.44)$$

where, for common macroscopic imaging,  $z_e = z_o - d_e$  is numerically negative, and  $\dot{z}_e = \dot{z}_o - \dot{d}_e$  is positive. Therefore, the sign of  $\alpha$  is opposite to the sign of  $\beta$ . This result implies that *the image plane must be rotated in the direction opposite to the direction of rotation of the object plane.*



**Figure 4.3** Object and image plane tilt (distances measured from principal planes). In the above cross-sectional (y-z plane) view, the object plane is tilted by an angle of  $\beta$  about the  $x$ -axis. The distances to the object and image plane pivots are specified from the respective principal planes.

---

We can further modify the relation Eq. (4.44) if the distances are specified with respect to Principal planes (Figure 4.3). Then,  $m_p$  represents the magnification between the image and object side principal planes and therefore, is equal to one; the ratio  $(\dot{z}_o - \dot{d}_e)/(z_o - d_e)$  represents the ratio of the distances of the principal-plane-to-object plane in the object side and principal-plane-to-image plane in the image side along the  $z$ -axis, and therefore is equal to the magnification along the  $z$ -axis. Letting  $\dot{u}_o = (\dot{z}_o - \dot{d}_e)$ ,  $u_o = (z_o - d_e)$ , and  $m_p = 1$ , we can rewrite Eq. (4.44) for the case if the image and object plane distances are measured from the Principal planes as:

$$\tan \alpha = \frac{\dot{u}_o}{u_o} \tan \beta \quad (4.45)$$

#### 4.2.3 Example: Focusing on a tilted object plane by tilting a lens using thin lens model

Determine the orientation (angle  $\alpha$ ) of a thin lens required to focus on a tilted object if the lens plane is not tilted.

1. Since we have a thin lens,  $m_p = 1$ ,  $R_\ell M_p R_\ell^T = I$ ,  $d_e = \dot{d}_e = 0$ .
2. Let  $p = \frac{\hat{n}_o(1)}{\hat{n}_o(3)}$  and  $q = \frac{\hat{n}_o(2)}{\hat{n}_o(3)}$ . Then,  $\hat{n}_o = [p \ q \ 1]^T$ .
3. Since the image plane is not rotated,  $\hat{n}_i = [0 \ 0 \ 1]^T$ .

The unknowns in this problem are the image plane distance  $\dot{z}_o$ , and the lens plane angle  $\alpha$ . In general, a rotation matrix has three degrees of freedom, however, since we have restricted the rotation of the lens plane to only about the  $x$ - and  $y$ - axes, the rotation matrix  $R_\ell$  in our problem has only two degrees of freedom. Further, we can describe the orientation of the lens plane using just the third column of  $R_\ell$  (the normal to the lens plane). Therefore, in total, we have three unknowns—the image plane distance and the two rotation angles of the lens plane. Another way to think about the number of knowns and unknowns is that we are required to determine the normal vector to the lens plane and the distance of the image plane. Since only two components of the normal vector are essential to determine the orientation of the plane (when the angles of rotation of the plane is restricted between  $\pm \frac{\pi}{2}$ ), we have three unknowns.

Substituting the known parameters in Eq. (4.23) we obtain:

$$-\frac{1}{z_o} \begin{bmatrix} p \\ q \\ 1 \end{bmatrix} + \frac{1}{\dot{z}_o} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \frac{r_{\ell,3}}{f} . \quad (4.46)$$

The three rows of the above equation yields:

$$\mathbf{r}_{\ell,3}(1) = -\frac{f}{z_o} p , \quad (4.47)$$

$$\mathbf{r}_{\ell,3}(2) = -\frac{f}{z_o} q , \quad (4.48)$$

and

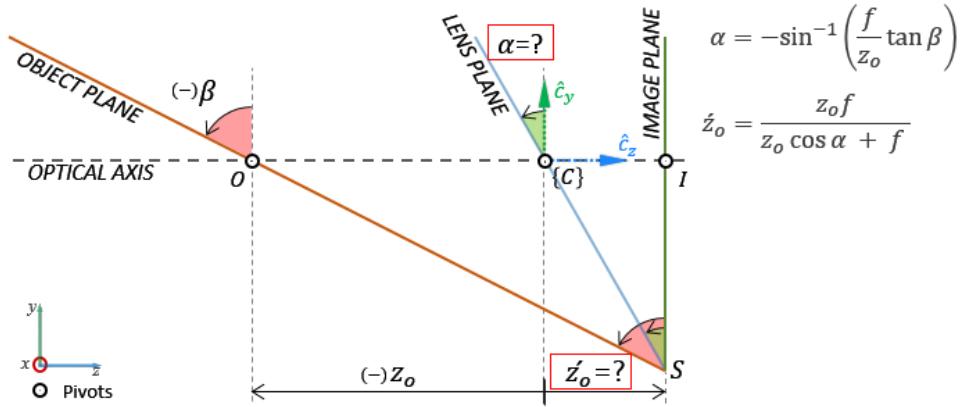
$$\dot{z}_o = -\frac{z_o f}{z_o \mathbf{r}_{\ell,3}(3) + f} . \quad (4.49)$$

To solve for the image plane distance  $\dot{z}_o$  in Eq. (4.49) we need  $\mathbf{r}_{\ell,3}(3)$ . Can we uniquely determine  $\mathbf{r}_{\ell,3}(3)$  from  $\mathbf{r}_{\ell,3}(1)$  and  $\mathbf{r}_{\ell,3}(2)$ ? A property of any rotation matrix is that each column (or row) has unit length. Therefore,  $\mathbf{r}_{\ell,3}(3) = \pm \sqrt{1 - (\mathbf{r}_{\ell,3}(1) + \mathbf{r}_{\ell,3}(2))^T}$ . Furthermore, if the rotation matrix is composed of elementary rotations only about the  $x$ - and  $y$ - axes, then  $\mathbf{r}_{\ell,3}(3)$ , which is a product of the cosine of the angles of rotations about the two axes, is guaranteed to be positive. Hence,

$$\mathbf{r}_{\ell,3}(3) = \sqrt{1 - (\mathbf{r}_{\ell,3}(1) + \mathbf{r}_{\ell,3}(2))^T} . \quad (4.50)$$

For a concrete example, suppose the object plane is tilted only about the  $x$ -axis by an angle  $\beta$ . Then, (as in the previous example),  $\hat{\mathbf{n}}_o = [0 \ -\sin \beta \ \cos \beta]^T$  and  $\hat{\mathbf{n}}_o = [0 \ -\tan \beta \ 1]^T$ . Therefore,  $\mathbf{r}_{\ell,3}(1) = 0$ ,  $\mathbf{r}_{\ell,3}(2) = \left(\frac{f}{z_o}\right) \tan \beta$ , and  $\mathbf{r}_{\ell,3}(3) = z_o^{-1} \sqrt{z_o^2 - f^2 \tan^2 \beta}$ . The image plane distance is given by  $\dot{z}_o = \frac{z_o f}{\sqrt{z_o^2 - f^2 \tan^2 \beta} + f}$ .

The dependence of  $\dot{z}_o$  on  $\beta$  (and implicitly on the amount of lens tilt as shown below in Eq. (4.52)) implies that we need to also translate the image plane location along the  $z$ -axis of  $\{\mathcal{C}\}$  to focus on a tilted plane. Instead, if we chose to focus on the tilted plane employing just sensor rotation, the distance of the sensor's pivot from the camera center remains fixed.



**Figure 4.4** Object and lens (thin lens model) plane tilt. In the above cross-sectional ( $y$ - $z$  plane) view, the object plane is tilted by an angle  $\beta$  about the  $x$ -axis at  $(0, 0, -z_o)$ . We would like to find the position of the image plane and orientation of the lens plane to focus on the tilted object surface.

Furthermore,  $\mathbf{r}_{\ell,3}(1) = 0$  implies that the lens is rotated only about the  $x$ -axis. If we let  $\mathbf{r}_{\ell,3} = [0 \ -\sin \alpha \ \cos \alpha]^T$  (the third column of  $R_\ell$ ), then the relationship between the lens plane rotation angle ( $\alpha$ ) and object plane rotation angle ( $\beta$ ) is obtained as (and shown in [Figure 4.4](#))

$$\alpha = -\sin^{-1}\left(\frac{f}{z_o} \tan \beta\right). \quad (4.51)$$

In the above equation  $z_o$  is numerically negative, therefore the sign of  $\alpha$  is same as the sign of  $\beta$ , which implies that the direction of rotation of the lens and object planes are congruent.

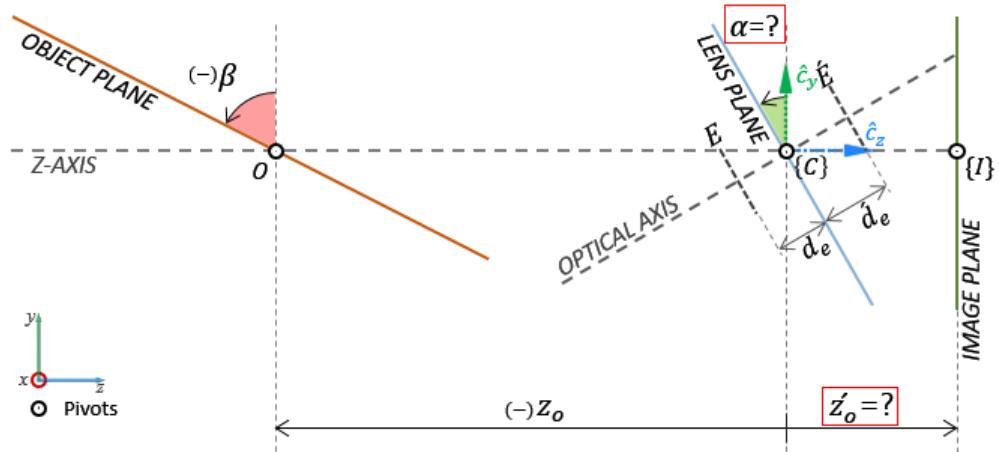
Furthermore, we obtain the image plane distance along the  $z$ -axis of  $\{C\}$  as:

$$z_o' = \frac{z_o f}{z_o \cos \alpha + f} \quad (4.52)$$

The sign of  $z_o'$  is positive, forming real and inverted image behind the lens, if  $|z_o| > f / \cos \alpha$ .

#### 4.2.4 Example: Focusing on a tilted object plane by tilting a lens using thick lens model

If the image plane is not rotated, what is the required angle of rotation of the lens to focus on an object surface that is titled about the  $x$ -axis by an angle  $\beta$  using a thick-lens model?



**Figure 4.5** Object and lens (thick lens model) plane tilt. In the above cross-sectional ( $y$ - $z$  plane) view, the object plane is tilted by an angle  $\beta$  about the  $x$ -axis at  $(0, 0, -z_o)$ . We would like to find the position of the image plane  $z'_o$  and orientation of the lens plane  $\alpha$  to focus on the tilted object surface.

The schematic of the problem is shown in [Figure 4.5](#). We can represent the orientation of the object plane that is tilted about the  $x$ -axis using the rotation matrix

$$R_o = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \beta & -\sin \beta \\ 0 & \sin \beta & \cos \beta \end{bmatrix}, \quad (4.53)$$

such that

$$\hat{\mathbf{n}}_o = \begin{bmatrix} 0 \\ -\tan \beta \\ 1 \end{bmatrix}. \quad (4.54)$$

In the previous problem (using thin lens model) we observed that the direction of rotation of the object- and lens-planes are congruent. Although we expect the exact angle of the lens plane to differ from the solution of the previous problem because of the thick lens model, there is no reason to suspect the direction of rotation of the lens to be different from the thin lens case. Therefore, the structure of the rotation matrix representing the lens plane's orientation is similar to that of the object plane:

$$R_\ell = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix}, \quad (4.55)$$

such that,

$$\mathbf{r}_{\ell,3} = \begin{bmatrix} 0 \\ -\sin \alpha \\ \cos \alpha \end{bmatrix}. \quad (4.56)$$

As the image plane is not tilted, we represent the normal of the image plane as:

$$\hat{\mathbf{n}}_i = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (4.57)$$

Therefore,  $\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3} = (\sin \alpha \tan \beta + \cos \alpha)$ ,  $\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3} = \cos \alpha$ , and

$$R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i = \begin{bmatrix} 0 \\ (1 - m_p) \sin \alpha \cos \alpha \\ m_p \cos^2 \alpha + \sin^2 \alpha \end{bmatrix}.$$

Substituting the above parameters into Eq. (4.23) we obtain:

$$\begin{aligned} & -\frac{1}{m_p[z_o - d_e(\sin \alpha \tan \beta + \cos \alpha)]} \begin{bmatrix} 0 \\ -\tan \beta \\ 1 \end{bmatrix} \\ & + \frac{1}{[\dot{z}_o - \dot{d}_e \cos \alpha]} \begin{bmatrix} 0 \\ (1 - m_p) \sin \alpha \cos \alpha \\ m_p \cos^2 \alpha + \sin^2 \alpha \end{bmatrix} = \frac{1}{f} \begin{bmatrix} 0 \\ -\sin \alpha \\ \cos \alpha \end{bmatrix} \end{aligned} \quad (4.58)$$

The third row, following simple algebraic steps, yields the formula for the image plane distance in terms of the angles of the object and lens planes as:

$$\boxed{\dot{z}_o = \dot{d}_e \cos \alpha + \frac{m_p f (m_p \cos^2 \alpha + \sin^2 \alpha) [z_o - d_e(\sin \alpha \tan \beta + \cos \alpha)]}{m_p \cos \alpha [z_o - d_e(\sin \alpha \tan \beta + \cos \alpha)] + f}} \quad (4.59)$$

From the second row we obtain:

$$\frac{\tan \beta}{m_p[z_o - d_e(\sin \alpha \tan \beta + \cos \alpha)]} + \frac{(1 - m_p)\sin \alpha \cos \alpha}{[\dot{z}_o - \dot{d}_e \cos \alpha]} = -\frac{\sin \alpha}{f}. \quad (4.60)$$

Substituting  $[\dot{z}_o - \dot{d}_e \cos \alpha]$  from Eq. (4.59) into Eq. (4.60) and writing  $\tan \beta = t$ ,  $\sin \alpha = s$ , and  $\cos \alpha = c$ , we obtain:

$$\frac{t}{m_p[z_o - d_e(st + c)]} + \frac{(1 - m_p)sc[m_p c\{z_o - d_e(st + c)\} + f]}{m_p f(m_p c^2 + s^2)[z_o - d_e(st + c)]} = -\frac{s}{f} \quad (4.61)$$

Multiplying by  $m_p[z_o - d_e(st + c)]$  we obtain:

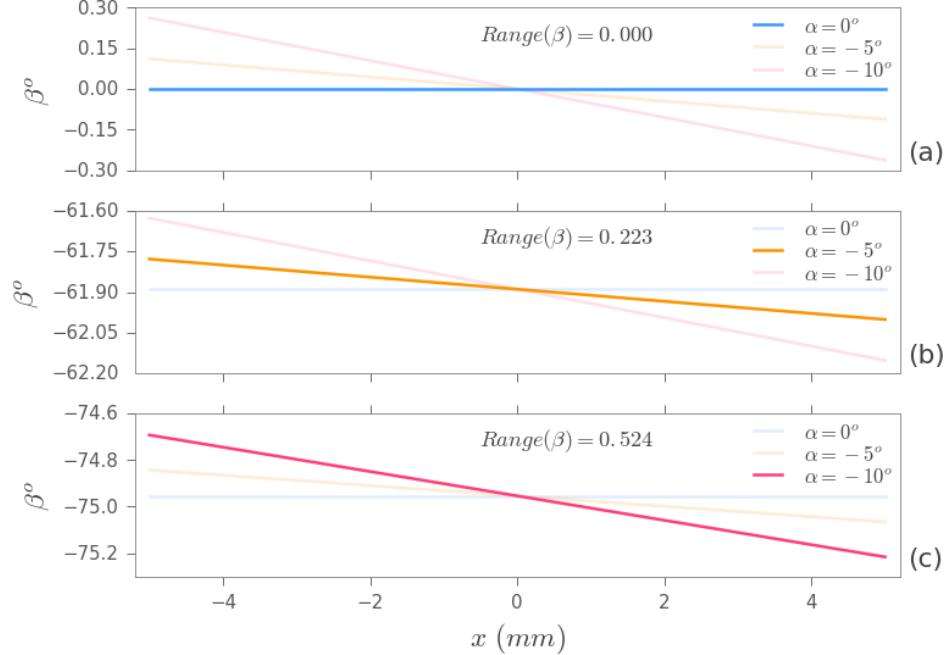
$$t + \frac{(1 - m_p)sc[m_p c\{z_o - d_e(st + c)\} + f]}{f(m_p c^2 + s^2)} = -\frac{sm_p z_o}{f} + \frac{sm_p d_e(st + c)}{f} \quad (4.62)$$

Following few algebraic steps, we obtain the finite conjugate imaging relationship between the object plane tilt angle  $\beta$  and lens plane tilt angle  $\alpha$  when the lens is rotated about a pivot (away from the entrance pupil) as:

$$\tan \beta = -\frac{\sin \alpha [m_p(z_o - d_e \cos \alpha) + f(1 - m_p)\cos \alpha]}{(f - m_p d_e \sin^2 \alpha)(m_p \cos^2 \alpha + \sin^2 \alpha) + m_p(1 - m_p)d_e \sin^2 \alpha \cos^2 \alpha} \quad (4.63)$$

In the discussion that follows we will use the notation  $g(\alpha, m_p, f, d_e, z_o)$  to represent the right-hand-side of the above equation. Eq. (4.63) is an implicit relationship between the angles  $\alpha$  and  $\beta$ . Comparing Eq. (4.63) with Eq. (4.51) immediately shows that for a given object plane tilt angle  $\beta$  the lens tilt angles obtained by the thick-lens (more accurate) model deviates from that obtained using a thin lens model. Further, the object plane angle (in focus) obtained via Eq. (4.63) for a given lens tilt angle  $\alpha$  depends on the location of the lens pivot point along the optical axis. The variation of  $\beta$  with respect to the pivot position (offset from entrance pupil,  $x$ ) for an object plane pivoted at  $z_e = -504 \text{ mm}$  from the entrance pupil of a  $f = 24 \text{ mm}$  lens with pupil

magnification  $m_p = 2.0$  is show in [Figure 4.6](#). The thin-lens model doesn't account for such deviations.



**Figure 4.6** Variation of  $\beta$  (y-axis) with respect to lens pivot position for (a)  $\alpha = 0^\circ$ , (b)  $\alpha = -5^\circ$ , and (c)  $\alpha = -10^\circ$ . The x-axis is the offset of the pivot position from the entrance pupil. The plots show that the range of  $\beta$  is large for larger values of  $\alpha$ . In each graph the two other plots are also plotted in lighter values for comparison of the slopes.

As the object plane distance  $z_o$  increases, the effective object-to-entrance-pupil distance ( $z_o - d_e \cos \alpha$ ) in Eq. (4.63) tends to a constant value for relatively small changes in the entrance-pupil-to-pivot-point distance  $d_e$ . Therefore, for relatively large object plane distances the variation of  $\beta$  is expected to be negligible for small changes in  $d_e$ .

#### ***Verification of formulae for focusing on a tilted object plane by tilting the lens***

Before we begin to examine the consequences of the focusing equations Eq. (4.59) and Eq. (4.63) that relates the lens plane orientation and image plane distance from the origin of camera frame  $\{C\}$  if the lens is tilted about a point away from the entrance pupil, it is imperative to ensure that the equations are verified. [Table 4.1](#) enumerates the results of our test. In order to test the equations, we implemented a  $f = 24.0$  mm thick lens model in Zemax using two paraxial surface—to ensure

aberration-free geometric imaging model—with pupil magnification  $m_p = 2$ . The lens surfaces were grouped within two coordinate break surfaces that allowed the lens to be tilted about a point  $d_e = -5.0 \text{ mm}$  away from the entrance pupil. The object plane surface was placed at  $z_o = -509.0 \text{ mm}$  from  $\{C\}$ . Then, for every object plane orientation  $\beta$  (col. 1), the appropriate lens tilt angle  $\alpha$  (col. 2) and image plane distance  $\dot{z}_o$  (col. 3) were obtained using Zemax’s optimization function, to minimize spot radius across the field. Following optimization for every  $\beta$ , the value of  $\alpha$  obtained from Zemax (along with the values of  $m_p, z_o, f$ ) was used to numerically compute  $\beta$  (col. 4) and  $\dot{z}_o$  (col. 5) using the derived equations Eq. (4.59) and Eq. (4.63). We can observe that the values of  $\beta$  and  $\dot{z}_o$  obtained numerically using the derived equations are very closely matched. It must be emphasized that the values ( $\alpha$  and  $\dot{z}_o$ ) obtained in Zemax were using ray-tracing and optimization, rather than numerical evaluation of analytic expressions.

**Table 4.1** Verification of imaging equations Eq. (4.59) and Eq. (4.63) for focusing on a tilted object plane by tilting a lens about a point away from the entrance pupil.

$\beta$ (Zemax) <sup>1</sup>	$\alpha$ (Zemax) <sup>2</sup>	$\dot{z}_o$ (Zemax) <sup>3</sup>	$\beta$ (numerical) <sup>4</sup>	$\dot{z}_o$ (numerical) <sup>5</sup>
0.0°	0.0°	24.17073 mm	2.07e-13°	24.17073 mm
-10.0°	-0.46989°	24.17163 mm	-9.99973°	24.17163 mm
25.0°	1.24260°	24.17701 mm	24.995702°	24.17701 mm
-40.0°	-2.23573°	24.19107 mm	-39.98214°	24.19107 mm
65.0°	5.70827°	24.30378 mm	64.91024°	24.30377 mm
-80.0°	-14.99585°	25.11194 mm	-79.74010°	25.11146 mm

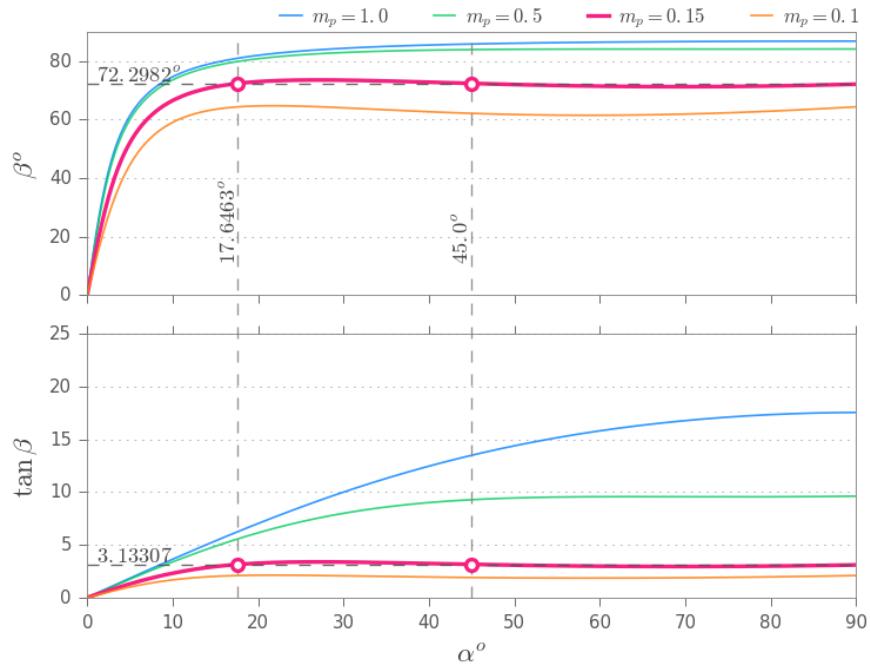
1. Object plane tilt  $\beta$  set in Zemax.
2. Lens plane tilt  $\alpha$  obtained using optimization using ray-tracing in Zemax.
3. Image plane distance  $\dot{z}_o$  obtained using optimization using ray-tracing in Zemax.
4. Object plane tilt  $\beta$  computed numerically using the value of  $\alpha$  in column 2.
5. Image plane distance  $\dot{z}_o$  computed numerically using the value of  $\alpha$  in column 2.

### **Consequences and analysis of the focusing equation**

Eq. (4.63) suggests that for a given lens tilt angle  $\alpha$ , we can determine the orientation of the plane-of-sharp focus  $\beta$ . But what if we need to compute  $\alpha$  given  $\beta$ ? Quite often we need the inverse relationship; that is, we need to determine the lens tilt angle required to focus on a tilted object plane. We will return to this question a few times in this section—first qualitatively in the imminent discussion followed by a detailed analysis near the end of this section. Another related question that we may ask at this point is whether the relationship between  $\alpha$  and  $\beta$ , as depicted using Eq. (4.63), always one-to-one? In other words, is the function  $g(\alpha, m_p, f, d_e, z_o)$ , which represents the right-hand-side of Eq. (4.63), *monotonic* and *invertible* within the interval  $-\pi/2 < \alpha < \pi/2$ ? If  $g(\alpha, m_p, f, d_e, z_o)$  is monotonic, then it follows that  $\beta = \tan^{-1}(g(\alpha, m_p, f, d_e, z_o))$  will be monotonic within  $-\pi/2 < \beta < \pi/2$  and we will have a one-to-one relationship between  $\alpha$  and  $\beta$ . A test to determine the monotonicity of  $g(\alpha, m_p, f, d_e, z_o)$  within an interval  $(a, b)$  is to examining if the first derivative,  $\dot{g}(\alpha, m_p, f, d_e, z_o)$ , changes sign within the interval. Later in this section, when we examine the case of rotating the lens about the entrance pupil, we will present a more detailed analysis of the first derivative of  $g(\alpha, m_p, f, d_e, z_o)$ . But first we carry out a qualitative analysis of Eq. (4.63) which relates the orientations of the object and lens planes if lens is pivoted about a point away from the entrance pupil. Although we will seldom rotate a lens about a point away from its entrance pupil because of the unwieldy distortions induced to the image field (see [Sec. 3.7.2](#)), the methods and insights developed in this study, nevertheless, will carry over to our examination of the latter case.

In [Figure 4.7](#) we have plotted values of  $\tan \beta$ —or  $g(\alpha, m_p, f, d_e, z_o)$ —and the object tilt angle  $\beta$  versus the lens tilt angle  $\alpha$  for four different values of  $m_p$  while keeping the parameters  $f (= 24\text{ mm})$ ,  $d_e (= -5\text{ mm})$  and  $z_o (= -509\text{ mm})$  fixed. We can observe that while the function  $g(\alpha, m_p, f, d_e, z_o)$  is monotonic for  $m_p = 0.5$  and  $m_p = 1.0$ , it becomes non-monotonic

for  $m_p = 0.15$  and  $m_p = 0.1$  as evidenced by the presence of stationary points (the exact locations are not important) in the corresponding plots. Consequently, for example, when  $m_p = 0.15$  we obtain the same value of  $\beta$  ( $\approx 72.3^\circ$ ) for two different values of  $\alpha$  ( $\approx 17.65^\circ$  and  $45.0^\circ$ ). These plots suggest that  $g(\alpha, m_p, f, d_e, z_o)$  is not always monotonic. In our simulations, we have found that  $g(\alpha, m_p, f, d_e, z_o)$  tends to become non-monotonic as the value of  $m_p$  becomes very small.



**Figure 4.7** Object plane angle  $\beta$  and  $\tan \beta = g(\alpha, m_p, f, d_e, z_o)$  versus lens tilt angle  $\alpha$  if a lens is rotated about a point away from the entrance pupil. In these plots  $z_o = -509 \text{ mm}$ ,  $f = 24 \text{ mm}$  and  $d_e = -5 \text{ mm}$ . The plots show that the function  $g(\alpha, m_p, f, d_e, z_o)$  is not always monotonic, especially for small values of  $m_p$ . For e.g., in the plots corresponding to  $m_p = 0.15$  the function yields the same value of  $\beta \approx 72.3^\circ$  for  $\alpha \approx 17.65^\circ$  and  $\alpha = 45^\circ$ .

---

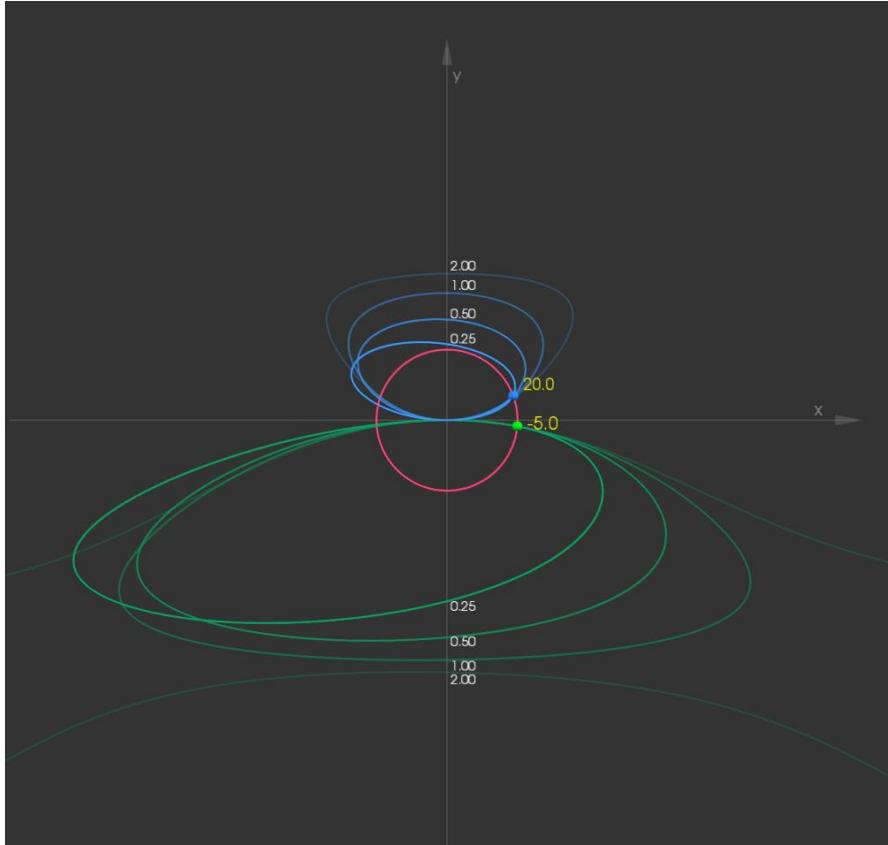
Apart from the monotonicity of  $g(\alpha, m_p, f, d_e, z_o)$ , we also need to consider the sign of the image plane distance from the exit pupil ( $\dot{z}_e$ ), equal to  $(\dot{z}_o - \dot{d}_e \cos \alpha)$ , obtained through Eq. (4.59). If  $|m_p \cos \alpha [z_o - d_e (\sin \alpha \tan \beta + \cos \alpha)]| < f$  in Eq. (4.59), then  $\dot{z}_e$  is negative which implies

that a *virtual image* is formed in front of the lens. Therefore, in order to form a *real image* on a sensor, the condition  $\dot{z}_e > 0$  must be satisfied.

Once the two independent conditions—the monotonicity of  $g(\alpha, m_p, f, d_e, z_o)$  and the formation of real image—are satisfied, we can return to the question regarding how to determine the lens tilt  $\alpha$  for a known object plane tilt angle  $\beta$ . Obtaining an expression for  $\alpha$  as a function of  $\beta$  from Eq. (4.63) is not straightforward. However, we can develop some insights into the problem if we substitute  $\cos \alpha = x$  and  $\sin \alpha = y$  into Eq. (4.63), which yields the implicit equation:

$$\begin{aligned} m_p d_e \tan \beta \textcolor{red}{y^4} - m_p (1 - 2m_p) d_e \tan \beta \textcolor{red}{x^2 y^2} - f m_p \tan \beta \textcolor{red}{x^2} - f \tan \beta \textcolor{red}{y^2} \\ + [m_p d_e - f(1 - m_p)] \textcolor{red}{xy} - z_o m_p \textcolor{red}{y} = 0. \end{aligned} \quad (4.64)$$

We can recognize Eq. (4.64) as a *quartic plane curve* of the form  $Ax^4 + By^4 + Cx^3y + Dx^2y^2 + Exy^3 + Fx^3 + Gy^3 + Hx^2y + Ixy^2 + Jx^2 + Ky^2 + Lxy + Mx + Ny + P = 0$ . The object plane tilt angle  $\beta$  along with the parameters  $z_o$ ,  $m_p$ ,  $f$  and  $d_e$  form the coefficients of this fourth degree plane curve. Further, since  $\cos \alpha = x$  and  $\sin \alpha = y$ , we also obtain a second curve—a unit circle with equation  $x^2 + y^2 = 1$ . The lens tilt angle  $\alpha$  (more precisely  $\cos \alpha$  and  $\sin \alpha$ ) satisfies both these equations, therefore it must be at the point of intersection of the two curves. In [Figure 4.8](#) we have plotted several quartic curves corresponding to different pupil magnifications  $m_p$  and two choices of lens rotations  $\alpha = +20^\circ$  and  $\alpha = -5^\circ$ . Note that the corresponding value of  $\beta$  for each curve in both groups will be different. The other parameters— $z_o$ ,  $d_e$  and  $f$ —are same for all curves. To keep the discussion simple, these parameters along with value the  $m_p$  for each curve used in the figure satisfy the conditions of monotonicity of  $g(\alpha, m_p, f, d_e, z_o)$  and real image. The curves in blue correspond to the  $\alpha = +20^\circ$  that satisfy Eq. (4.64), and the green curves correspond to  $\alpha = -5^\circ$ . Since the constant in Eq. (4.64)—corresponding to  $P$  in the general quartic equation—is zero, all curves pass through the origin. Additionally, since the coefficient of the  $x$  term is zero while the coefficient of the  $y$  term is non-zero, the curves are either above the  $x$ -axis



**Figure 4.8** Determination of lens tilt angle  $\alpha$  for known object tilt angle  $\beta$  using point of intersection of *quartic plane curve* with the unit circle. If a lens is rotated about a point away from the entrance pupil, the equation for  $\beta$  as a function of  $\alpha, m_p, z_o, f$  and  $d_e$  can be expressed as a quartic plane curve in Cartesian coordinates with  $x = \cos \alpha, y = \sin \alpha$ . Furthermore, the intersection of the quartic curve with the unit circle yields the unknown lens tilt angle  $\alpha$ . The figure plots two groups of quartic curves—the blue curves correspond to equations for different values of  $m_p$  (0.25, 0.5, 1.0 and 2.0) and  $\beta$  but the same  $\alpha = +20^\circ$  while the green curves corresponds to  $\alpha = -5^\circ$ . For all curves  $z_o = -509\text{ mm}$ ,  $f = 24\text{ mm}$  and  $d_e = -5\text{ mm}$ .

---

(for positive values of  $\alpha$ ) or below the y-axis (for negative values of  $\alpha$ ). As shown in the figure, all such quartic curves (for which the parameters satisfy the monotonicity condition) intersect the unit circle at two points—in the first and second quadrants if  $\alpha > 0$ , or in the third and fourth quadrants if  $\alpha < 0$ . However, since  $x = \cos \alpha$  in Eq. (4.64) the abscissa of the point of intersection must always be positive for  $-\pi/2 < \alpha < \pi/2$ . Consequently, we can determine a unique point

of intersection of the quartic curve (given  $\beta$ ,  $m_p$ ,  $f$ ,  $d_e$  and  $z_o$ ) with the unit circle that correspond to the unknown lens tilt angle  $\alpha$ .

Based on the above discussion we see that it is possible to implement an algorithm to find the point of intersection of a quartic curve and the unit circle—for example, using Newton’s method—to determine  $\alpha$  given  $\beta$ . Alternatively, if a good initial estimate of  $\alpha$  as a starting point is known, an iterative algorithm that converges towards the true point of intersection along the unit circle may be used to find the  $\alpha$ . We demonstrate such an algorithm at the end of this section.

Rotating the lens about a point away from the entrance pupil distorts the image field in complex ways and induces parallax between corresponding scene points in sequence of images obtained under lens rotations. Therefore, in several imaging applications that require multiple image captures under lens rotation it is imperative to rotate the lens about the entrance pupil. In the rest of this section we will undertake a closer examination of this scenario.

We can obtain the equations for image plane distance  $\acute{z}_o$  and plane of sharp focus orientation  $\beta$  for the case when the lens is pivoted at the entrance pupil by substituting  $d_e = 0$  and  $\acute{d}_e = d$  (where,  $d$  is the distance of the exit pupil from the entrance pupil) in Eqs. (4.59) and (4.63) respectively yielding:

$$\boxed{\acute{z}_o = d \cos \alpha + \frac{m_p z_o f (m_p \cos^2 \alpha + \sin^2 \alpha)}{m_p z_o \cos \alpha + f}}, \quad (4.65)$$

and

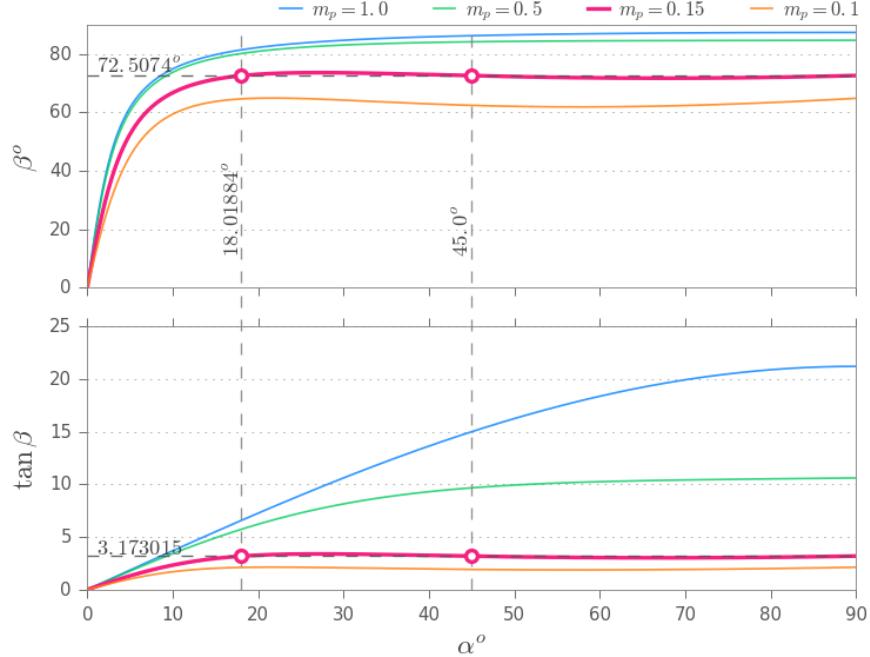
$$\boxed{\tan \beta = -\frac{\sin \alpha [m_p z_o + f(1 - m_p) \cos \alpha]}{f(m_p \cos^2 \alpha + \sin^2 \alpha)}}. \quad (4.66)$$

**Table 4.2** Verification of imaging equations Eq. (4.65) and Eq. (4.66) for focusing on a tilted object plane by tilting a lens about the entrance pupil.

$\beta$ (Zemax) <sup>1</sup>	$\alpha$ (Zemax) <sup>2</sup>	$z'_o$ (Zemax) <sup>3</sup>	$\beta$ (numerical) <sup>4</sup>	$z'_o$ (numerical) <sup>5</sup>
0.0°	0.0°	29.17073 mm	-2.2e-15°	29.17073 mm
-10.0°	-0.46989°	29.17145 mm	-10.0°	29.17145 mm
25.0°	1.24249°	29.17572 mm	25.0°	29.17572 mm
-40.0°	-2.23504°	29.18687 mm	-40.0°	29.18687 mm
65.0°	5.69682°	29.27607 mm	65.0°	29.27607 mm
-80.0°	-14.79587°	29.90304 mm	-80.0°	29.90304 mm

1. Object plane tilt  $\beta$  set in Zemax.
2. Lens plane tilt  $\alpha$  obtained using optimization using ray-tracing in Zemax.
3. Image plane distance  $z'_o$  obtained using optimization using ray-tracing in Zemax.
4. Object plane tilt  $\beta$  computed numerically using the value of  $\alpha$  in column 2.
5. Image plane distance  $z'_o$  computed numerically using the value of  $\alpha$  in column 2.

[Table 4.2](#) enumerates the results of verifying the above equations against Zemax. The optical system used in the test is like that used to verify the Scheimpflug imaging equations for lens rotation about a point away from the entrance pupil, except that since the pivot was shifted to the entrance pupil, the object distance  $z_o$  changed from  $-509.0\text{ mm}$  to  $-504.0\text{ mm}$ . The table shows that the numerically computed value of object plane tilt angle  $\beta$  using the analytic expressions matches exactly to that used in Zemax to tilt the object plane. Also, the value of  $z'_o$  obtained in Zemax using ray tracing and optimization is in exact match with that obtained using the analytic expression we derived, proving the accuracy of the above expressions.



**Figure 4.9** Object plane angle  $\beta$  and  $\tan \beta = g(\alpha, m_p, f, z_o)$  versus lens tilt angle  $\alpha$  if a lens is rotated about the entrance pupil. The plots show that the function  $g(\alpha, m_p, f, z_o)$  is not always monotonic, especially for small values of  $m_p$ . For e.g., in the plots corresponding to  $m_p = 0.15$  the function yields the same value of  $\beta \approx 72.5^\circ$  for  $\alpha \approx 18.02^\circ$  and  $\alpha = 45^\circ$ .  $z_o = -509$  mm and  $f = 24$  mm for all plots.

---

Comparing Eqs. (4.65) and (4.66) with Eqs. (4.59) and (4.63) respectively, we can immediately observe that the equations are far less complex when the lens is rotated about the entrance pupil. However, just like before, the function  $g(\alpha, m_p, f, z_o)$  that represents the right-hand-side of Eq. (4.66) is not always monotonic. As shown in [Figure 4.9](#),  $g(\alpha, m_p, f, z_o)$  is monotonic for  $m_p = 1.0$  and  $m_p = 0.5$  but non-monotonic for  $m_p = 0.15$  and  $m_p = 0.1$ .

The following two subsections are devoted to a deeper examination of the condition of monotonicity of  $g(\alpha, m_p, f, z_o)$  and the development of an iterative algorithm for determining the lens tilt angle  $\alpha$ .

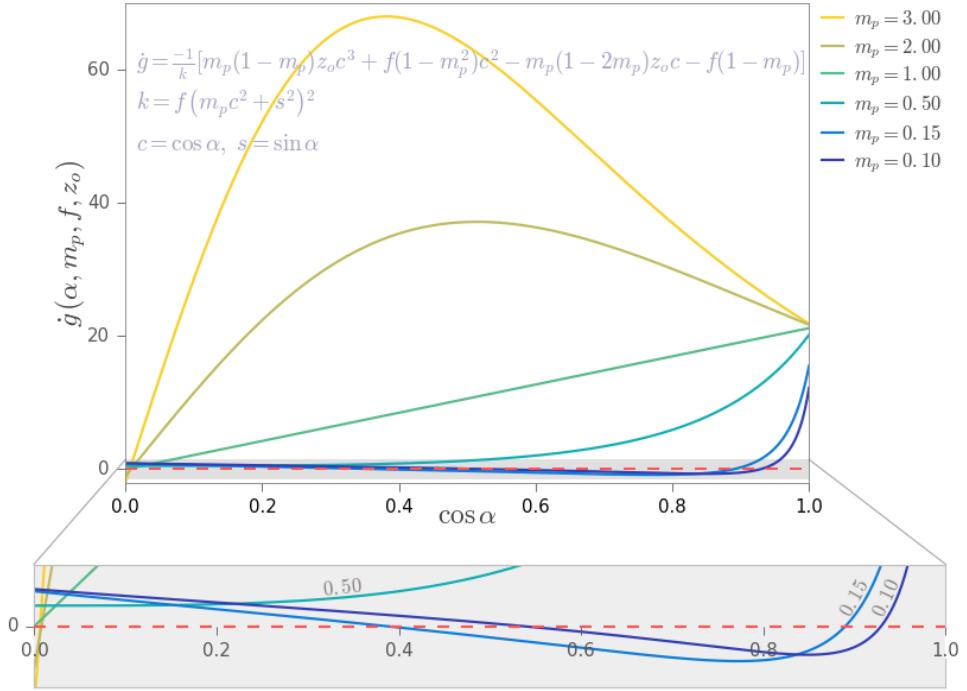
### Condition for monotonicity of $g(\alpha, m_p, f, z_o)$

The first derivative of  $g(\alpha, m_p, f, z_o)$  with respect to  $\alpha$  is:

$$\begin{aligned}\dot{g} = \frac{-1}{k} & [m_p(1 - m_p)z_o \cos^3 \alpha + f(1 - m_p^2) \cos^2 \alpha - m_p(1 - 2m_p)z_o \cos \alpha \\ & - f(1 - m_p)],\end{aligned}\quad (4.67)$$

where,  $k = f(m_p \cos^2 \alpha + \sin^2 \alpha)^2 = f[1 - (1 - m_p) \cos^2 \alpha]^2$ .

We observe that the first derivative  $\dot{g}(\alpha, m_p, f, z_o)$  is a cubic equation in  $\cos \alpha$ . A cubic equation has at least one real root which implies that the graph of  $\dot{g}(\alpha, m_p, f, z_o)$  must cross the real x-axis at least once. However, in our problem, the lens rotation angle  $\alpha$  is restricted to within  $\pm\pi/2$ . Consequently, we are only concerned with those roots of  $\dot{g}(\alpha, m_p, f, z_o)$  that are in the open interval  $(0, 1)$ , because  $0 < \cos \alpha < 1$  for  $-\pi/2 < \alpha < \pi/2$ . More concretely, if  $\dot{g}(\alpha, m_p, f, z_o)$  has a real positive root in the interval  $(0, 1)$  implying that it changes sign, then  $g(\alpha, m_p, f, z_o)$  is non-monotonic within  $-\pi/2 < \alpha < \pi/2$ . In such a case, we cannot find a unique  $\alpha$  for a given value of  $\beta$ . In [Figure 4.10](#) we have plotted the first derivative  $\dot{g}(\alpha, m_p, f, z_o)$  for varying values of  $m_p$ . We can see that the plots of  $\dot{g}(\alpha, m_p, f, z_o)$  for  $m_p = 0.15$  and  $m_p = 0.10$  have at least two roots as they cross the real x-axis twice with the interval  $(0, 1)$ . Of course, this result was expected as we have already seen in [Figure 4.10](#) that  $g(\alpha, m_p, f, z_o)$  is non-monotonic for  $m_p = 0.15$  and  $m_p = 0.10$ .



**Figure 4.10** Plots of the first derivative of  $g(\alpha, m_p, f, z_o)$ . The plots first derivative for  $m_p = 0.15$  and  $m_p = 0.10$  cross the real x-axis twice within the interval  $0 < \cos \alpha < 1$  implying real roots in the interval  $(0, 1)$ . Consequently,  $g(\alpha, m_p, f, z_o)$  is non-monotonic in  $-\pi/2 < \alpha < \pi/2$ .

Additionally, we can easily examine the first derivative when  $m_p = 1$ , which is given as

$$\dot{g}(\alpha, m_p = 1, f, z_o) = -\frac{z_o}{f} \cos \alpha . \quad (4.68)$$

For  $m_p = 1$ , the first derivative is a linear function of  $\cos \alpha$  and crosses the  $x$ -axis at the origin as can be seen in [Figure 4.10](#). Therefore, it has no real roots in the open interval  $(0, 1)$ , implying that  $g(\alpha, m_p = 1, f, z_o)$  is monotonic. Furthermore, since  $z_o$  is numerically negative (directed distance),  $\dot{g}(\alpha, m_p = 1, f, z_o)$  is a monotonically increasing function implying that  $g(\alpha, m_p, f, z_o)$ , and consequently  $\beta$ , increases with  $\alpha$ .

Heretofore we have used visualizations to analyze the conditions under which it is possible to invert the function  $g(\alpha, m_p, f, z_o)$ . We now proceed to find an analytic expression that can be

used to test the monotonicity of  $(\alpha, m_p, f, z_o)$ . We can, of course, use any numerical computation tool to find the cubic roots of the first derivative and verify if the first derivative has real roots in the open interval  $(0, 1)$ . In fact, our derivation is based on the algebraic method for solving cubic roots published by Gerolamo Cardano in his treatise *Ars Magna* in 1545 [81].

The roots of a cubic polynomial  $ax^3 + bx^2 + cx + d = 0$  is given as:

$$\begin{aligned} z_1 &= -\frac{b}{3a} + \sqrt[3]{R + \sqrt{D}} + \sqrt[3]{R - \sqrt{D}}, \\ z_2 &= -\frac{b}{3a} - \frac{1}{2} \left( \sqrt[3]{R + \sqrt{D}} + \sqrt[3]{R - \sqrt{D}} \right) + \frac{1}{2} i\sqrt{3} \left( \sqrt[3]{R + \sqrt{D}} - \sqrt[3]{R - \sqrt{D}} \right), \\ z_3 &= -\frac{b}{3a} - \frac{1}{2} \left( \sqrt[3]{R + \sqrt{D}} + \sqrt[3]{R - \sqrt{D}} \right) - \frac{1}{2} i\sqrt{3} \left( \sqrt[3]{R + \sqrt{D}} - \sqrt[3]{R - \sqrt{D}} \right), \end{aligned} \quad (4.69)$$

where,

$$D = Q^3 + R^2, \quad (4.70)$$

and

$$\begin{aligned} Q &:= \frac{c}{3a} - \frac{b^2}{9a^2}, \\ R &:= \left( -\frac{b^3}{27a^3} + \frac{bc}{6a^2} - \frac{d}{2a} \right). \end{aligned} \quad (4.71)$$

We can determine the nature of the roots from  $D$ —if  $D > 0$ , then  $R \pm \sqrt{D}$  is real resulting in one real root ( $z_1$ ) and two roots that are complex conjugates ( $z_2, z_3$ ); if  $D = 0$ , then the imaginary terms of  $z_2$  and  $z_3$  vanishes resulting in all three roots being real and at least two equal; and if  $D < 0$ , then  $\sqrt[3]{R + \sqrt{D}}$  and  $\sqrt[3]{R - \sqrt{D}}$  becomes complex conjugates resulting in all three roots being real.

Since  $k$  in first derivative represented by Eq. (4.67) is positive in our problem, the roots of the first derivative are same as the roots of the scaled cubic equation

$$x^3 + \frac{f(1+m_p)}{m_p z_o} x^2 - \frac{(1-2m_p)}{(1-m_p)} x - \frac{f}{m_p z_o} = 0, \quad (4.72)$$

with coefficients  $a = 1$ ,  $b = f(1+m_p)/(m_p z_o)$ ,  $c = -(1-2m_p)/(1-m_p)$ , and  $d = -f/(m_p z_o)$  and  $x = \cos \alpha$ .

Eq. (4.72) is not suitable if  $m_p = 1$ ; however, we have already seen that when  $m_p = 1$ , the first derivative represents a line passing through the origin. For  $m_p > 1$ , the coefficient of the linear term— $c$ —is larger than both the constant and quadratic terms. In fact, the magnitude of  $c$  tends to two, starting from a very large magnitude, as  $m_p$  increases from one. At the same time the magnitude of  $d$  tends towards zero. Therefore, for  $m_p \geq 1$  the curve represented by Eq. (4.72) never crosses the real x-axis in the interval  $(0 + \Delta_x, 1)$ ; where  $\Delta_x$  is a very small number, whose exact value depends on  $d$  and corresponds to an angle  $\alpha$  that is very close to  $\pm\pi/2$ . Therefore, we only need to test for the monotonicity of  $g(\alpha, m_p, f, z_o)$  if  $m_p < 1$ .

Based on the discussion of nature of the cubic roots, we would expect that if all three roots are real, then there is a high chance that one or more of these real valued roots would lie within the interval  $(0, 1)$ . Indeed, based on tens of thousands of randomly generated combinations of parameters  $m_p$ ,  $f$  and  $z_o$  we have found that the only instances in which we obtain real roots of Eq. (4.72) in the interval  $(0 + \Delta_x, 1)$  is when  $D < 0$ . Finally, substituting the expressions for the coefficients  $a$ ,  $b$ ,  $c$ , and  $d$  from Eq. (4.72) into Eq. (4.71) and Eq. (4.70) we get the *sufficient condition* for  $g(\alpha, m_p, f, z_o)$  to be monotonic and invertible if  $m_p < 1$  as:

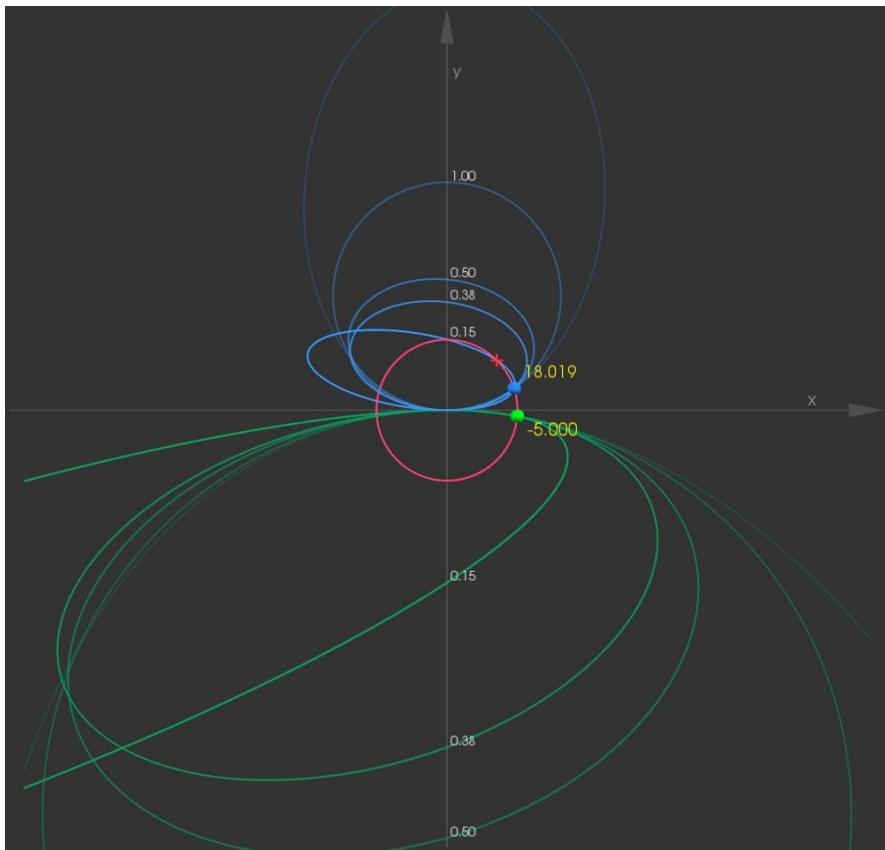
$$\left[ \frac{f}{2m_p z_o} - \frac{f(1+m_p)(1-3m_p)}{6(1-m_p)m_p z_o} - \frac{f^3(1+m_p)^3}{27m_p^3 z_o^3} \right]^2 - \left[ \frac{(1-2m_p)}{3(1-m_p)} + \frac{f^2(1+m_p)^2}{9m_p^2 z_o^2} \right]^3 > 0 \quad (4.73)$$

### **Algorithm for finding $\alpha$ for known $\beta$**

The Eq. (4.66) can be used to easily determine the tilted object plane orientation  $\beta$  that is brought to focus if the lens is tilted by an angle  $\alpha$  about the center of the entrance pupil. However, it is often required to find the lens tilt angle  $\alpha$  as a function of a known object plane tilt angle  $\beta$ . Deriving a closed-form inverse relation of the Eq. (4.66) is not easy. Therefore, we will develop an iterative algorithm for determining  $\alpha$ . The central idea behind the algorithm was introduced earlier when we saw that the general relationship between  $\alpha$  and  $\beta$ , if the lens is pivoted about a point away from the entrance pupil, lead to a quartic plane curve equation parameterized by  $\beta$ ,  $z_o$ ,  $f$  and  $d_e$ . The point of intersection of the quartic curve with the unit-circle (and having a positive value of abscissa) yielded  $\alpha$ . In a similar vein, by substituting  $x = \cos \alpha$  and  $y = \sin \alpha$  in Eq. (4.66), we obtain the following implicit equation if the lens is rotated about the entrance pupil:

$$fm_p \tan \beta \textcolor{red}{x^2} + f \tan \beta \textcolor{red}{y^2} + f(1 - m_p) \textcolor{red}{xy} + m_p z_o \textcolor{red}{y} = 0 \quad (4.74)$$

Similar to our earlier observation on the reduction of complexity of the equation relating  $\alpha$  and  $\beta$ , the implicit equation too reduces from a fourth-degree *quartic plane curve* (Eq. (4.64)) to a second-degree *quadratic plane curve*. We have plotted several such curves in two groups in [Figure 4.11](#). The curves in each group, distinguished by the green and blue lines, belong to  $\alpha = -5^\circ$  and  $\alpha \approx 18.02^\circ$  respectively. In each group the various curves correspond to different values of  $m_p$  (and  $\beta$ ). For all curves in the figure the focal length  $f = 24 \text{ mm}$ , and the object plane distance  $z_o = -509 \text{ mm}$ . The shapes of these curves are almost always elliptic, although for small values of  $m_p$  we have also observed parabolic and hyperbolic shapes. At  $m_p = 1$  the curve, irrespective of other parameters, is a circle. The point of intersection of each quadratic curve with the unit circle in the first or third quadrants yields the lens tilt angle  $\alpha$  for positive or negative sign of object tilt angle  $\beta$  respectively. The red cross in the figure depicts the second possible solution for  $\alpha$  (equal to  $45.0^\circ$ )



**Figure 4.11** Determination of lens tilt  $\alpha$  for known object plane tilt  $\beta$  using point of intersection of *quadratic plane curve* with the unit circle. If a lens is rotated about the entrance pupil, the equation relating  $\beta$  as a function of  $\alpha$ ,  $m_p$ ,  $z_o$ , and  $f$  can be expressed as a quadratic plane curve with  $x = \cos \alpha$ ,  $y = \sin \alpha$ . The intersection of this curve with the unit circle yields the unknown  $\alpha$ . The figure plots two groups of curves—the blue curves correspond to equations for different values of  $m_p$  (0.15, 0.38, 0.5, 1.0 and 2.0) and  $\beta$  but the same  $\alpha = +18.019^\circ$  while the green curves corresponds to  $\alpha = -5^\circ$ . The red cross indicates a second point of intersection of one of the curves with the unit circle implying that there exist two values of  $\alpha$  ( $18.019^\circ$  and  $45^\circ$ ) that evaluates to the same value of  $\beta$  using Eq. (4.66). Therefore, the relationship between  $\alpha$  and  $\beta$  is not unique for  $m_p = 0.15$  for choice of object distance  $z_o$  (-509 mm) and focal length  $f$  (24 mm).

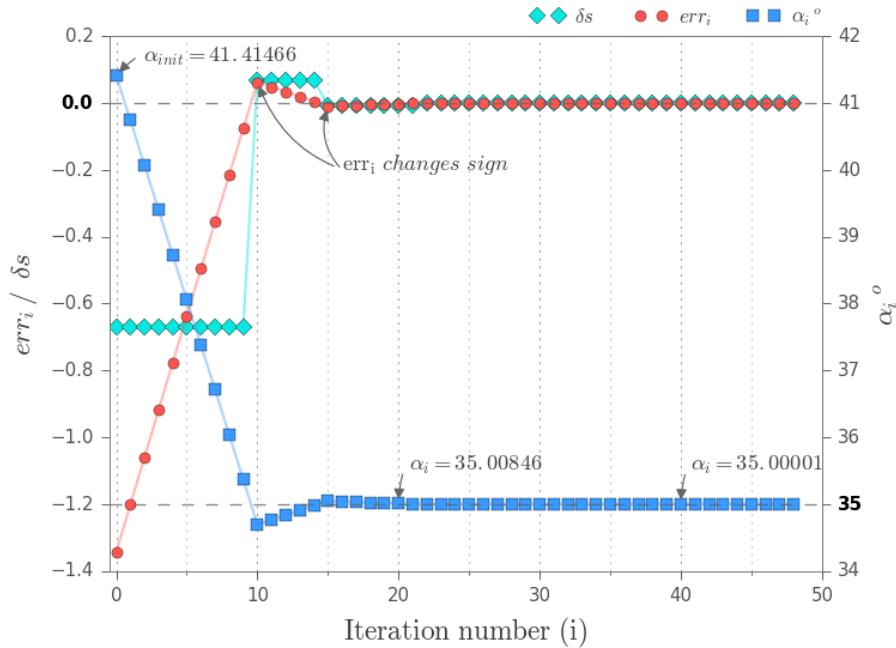
---

when  $m_p = 0.15$ , implying that the system does not meet the *sufficient condition* for determining  $\alpha$ .

As discussed previously, one possible method to find the lens tilt angle  $\alpha$  for known  $\beta$ ,  $m_p$ ,  $f$ , and  $z_o$  would be to numerically compute the points of intersection of the corresponding quadratic curve with the unit circle, for example using Newton's method, followed by selecting the point of

intersection from the appropriate quadrant depending on the sign of  $\beta$ . In our problem, since the desired lens tilt angle is constrained to reside on the unit circle, instead of finding the roots of a cubic equation, we can use an iterative algorithm to find  $\alpha$ , provided we can find a good starting point that is relatively close to the desired  $\alpha$ . In fact, we do have a good starting point—the value of  $\alpha$  for  $m_p = 1$ . If  $m_p = 1$  and the lens is rotated about the entrance pupil, the equation relating  $\alpha$  and  $\beta$  is equivalent to rotating a thin lens represented by Eq.(4.51).

The iterative algorithm used for determining the lens tilt angle  $\alpha$  for a known value of  $\beta$  is shown in [Table 4.3](#). In addition, [Figure 4.12](#) and [Figure 4.13](#) visually explain how the algorithm functions. [Figure 4.12](#) describes how  $err_i$ ,  $\delta s$ , and  $\alpha_i$  evolve over iteration number as the algorithm converges to the value of  $\alpha$  starting from an initial estimate given by thin-lens approximation. The figure also shows how the step size  $\delta s$  is modified when the  $i^{th}$  error  $err_i$  changes sign. While [Figure 4.12](#) illustrates how the algorithm converges to the right values of  $\alpha$ , [Figure 4.13](#) shows how  $\alpha$  evolves along the unit circle for two different choices of  $m_p$ . Both figures demonstrate that for typical optical systems the algorithm converges rapidly to the true value of the  $\alpha$ . Furthermore, more number of iterations improves the precision of the estimate. It is important to note that the system with  $m_p = 0.12$  (green) is only used to demonstrate the algorithm. Although the parameters meet the sufficient condition test, the image plane of this hypothetical system is located several meters behind the exit pupil. On the other hand, an optical system with  $m_p$  around 1.5 is quite common. For both curves,  $z_o = -509\text{ mm}$  and  $f = 50\text{ mm}$ .



**Figure 4.12** Example determination of lens tilt angle  $\alpha$ . The plot shows the iteration variables  $err_i$  (red),  $\delta s$  (cyan), and  $\alpha_i$  (blue) against the iteration number. While the range of values of  $err_i$  and  $\delta s$  are shown on the left vertical axis, the right vertical axis represents  $\alpha_i$ . The true value of  $\alpha$  is  $35^\circ$ , which corresponds to  $\beta = 81.55^\circ$  with  $m_p = 0.12$ ,  $z_o = -509\text{ mm}$  and  $f = 24.0\text{ mm}$ . The algorithm starts with an initial estimate of  $\alpha \approx 41.42^\circ$  (obtained using thin lens approx.) and converges to 35.008 in the first 20 iterations. The final value of  $\alpha$ , after 48 iterations, is 35.00001. The plot also shows that the step size  $\delta s$  remains constant as the error in each iteration decreases until the value of the error changes sign; at which point the step's value and sign are modified.

---

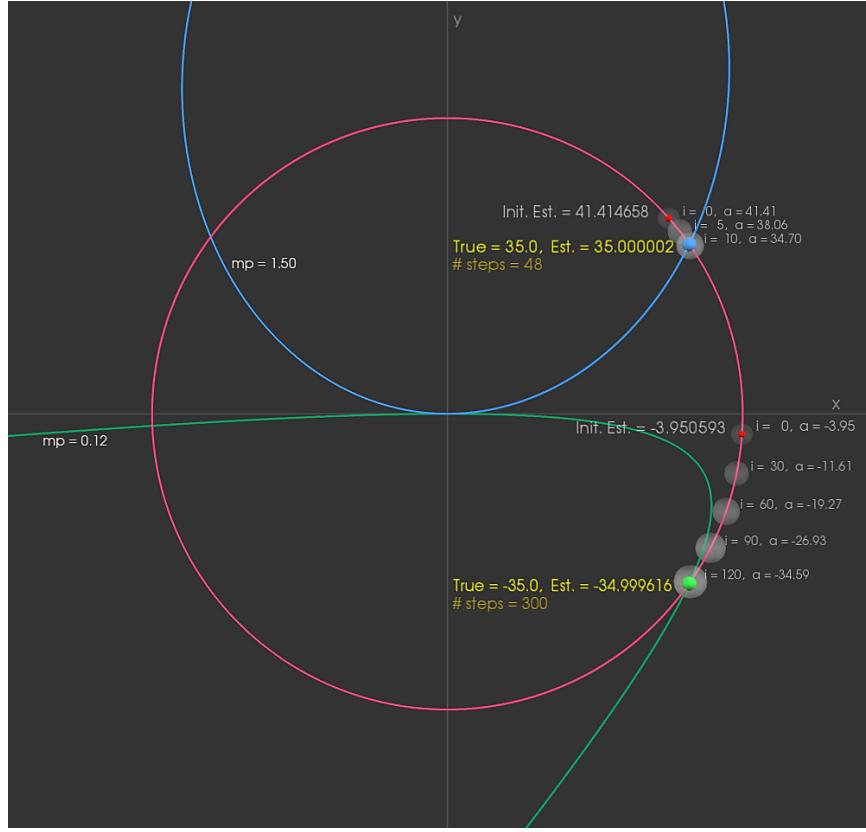
**Table 4.3** Algorithm for finding lens tilt  $\alpha$  required to focus on an object plane tilted by  $\beta$ .

Objective

Given object plane tilt angle  $\beta$ , determine  $\alpha$ , the lens tilt angle about the entrance pupil, required to focus on object plane that is at a distance  $z_o$  from the lens pivot. The lens has focal length  $f$  and pupil magnification  $m_p$ . The iteration quits if either the specified accuracy for the result,  $err_{max}$ , or the specified maximum steps,  $N$ , is reached.

Algorithm

1. Define function:  $g(\alpha, m_p, f, z_o) := -\frac{\sin \alpha [m_p z_o - f(1-m_p) \cos \alpha]}{f[m_p \cos^2 \alpha + \sin^2 \alpha]}$
2. Compute initial estimate of  $\alpha$  to provide a starting point:  $\alpha_{init} = -\sin^{-1}\left(\frac{f}{z_o} \tan \beta\right)$
3. Compute initial error:  $err_{init} = \tan \beta - g(\alpha_{init}, m_p, f, z_o)$
4. Compute initial value of step to increment/ decrement  $\alpha_i$ :  $\delta s = \frac{1}{2} err_{init}$
5. Set iteration variables for error:  $err_i = err_{init}$ ,  $err_{last} = 0$
6. Set iteration variable for angle  $\alpha$ :  $\alpha_i = \alpha_{init}$
7. Set step counter:  $N = 0$
8. *while* ( $|err_i| > err_{max}$  or  $N < N_{max}$ ) repeat:
  - a. Increment step counter:  $N = N + 1$
  - b. Update estimate of lens tilt angle  $\alpha$ :  $\alpha_i = \alpha_i + \delta s$
  - c. Store current error:  $err_{last} = err_i$
  - d. Compute new error:  $err_i = \tan \beta - g(\alpha_i, m_p, f, z_o)$
  - e. Modify step direction and size if error changes sign: *if*  $err_i * err_{last} < 0$ :
    - o Set new step size:  $\delta s = \frac{1}{2} (err_i - err_{last})$



**Figure 4.13** Inner workings of the iterative algorithm for determining  $\alpha$  given  $\beta$ . The ellipse (blue) corresponds to  $m_p = 1.5$ ,  $\beta = 81.55^\circ$ . The hyperbola (green) corresponds to  $m_p = 0.12$ ,  $\beta = -35.04^\circ$ . The true values of  $\alpha$  for the blue and green curves are  $\alpha = 35^\circ$  and  $\alpha = -35^\circ$  respectively, shown by the blue and green spheres on the unit circle (magenta). The initial estimates for  $\alpha$  are shown by the small red spheres. The opacity, position, and numeric tags of white spheres depict the direction of convergence, the value of  $\alpha_i$  in the  $i^{th}$  iteration and the iteration number. For  $m_p = 1.5$ , the algorithm converges at a faster rate than for  $m_p = 0.12$  as indicated by the number of steps and precision of the estimated  $\alpha$  in the two cases. It is important to note that the system with  $m_p = 0.12$  (green) is only used to demonstrate the algorithm.

### 4.3 Summary

We derived the formula to create a geometrically sharp image of a tilted object plane (described by the normal  $\hat{\mathbf{n}}_o$ ) on a tilted image plane (described by the normal  $\hat{\mathbf{n}}_i$ ) through a tilted lens (described by the rotation matrix  $R_\ell$ ) as (Eq. (4.23)):

$$-\frac{\hat{\mathbf{n}}_o}{m_p[z_o - d_e(\hat{\mathbf{n}}_o^T \mathbf{r}_{\ell,3})]} + \frac{R_\ell M_p R_\ell^T \hat{\mathbf{n}}_i}{[\hat{z}_o - \hat{d}_e(\hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})]} = \frac{\mathbf{r}_{\ell,3}}{f}.$$

We saw that the above equation is most general and special configurations of both frontoparallel and Scheimpflug imaging readily fall out of it.

Using the above formula, we derived a set of formula for focusing on a tilted object plane, if lens' pivot is away from the entrance pupil as (Eq. (4.59)):

$$\dot{z}_o = \hat{d}_e \cos \alpha + \frac{m_p f (m_p \cos^2 \alpha + \sin^2 \alpha) [z_o - d_e (\sin \alpha \tan \beta + \cos \alpha)]}{m_p \cos \alpha [z_o - d_e (\sin \alpha \tan \beta + \cos \alpha)] + f},$$

and (Eq. (4.63))

$$\tan \beta = -\frac{\sin \alpha [m_p (z_o - d_e \cos \alpha) + f(1 - m_p) \cos \alpha]}{(f - m_p d_e \sin^2 \alpha)(m_p \cos^2 \alpha + \sin^2 \alpha) + m_p (1 - m_p) d_e \sin^2 \alpha \cos^2 \alpha}.$$

If, however the lens is pivoted at the center of the entrance pupil, these formulae simplify to (Eq. (4.65)):

$$\dot{z}_o = d \cos \alpha + \frac{m_p z_o f (m_p \cos^2 \alpha + \sin^2 \alpha)}{m_p z_o \cos \alpha + f},$$

and (Eq. (4.66))

$$\tan \beta = -\frac{\sin \alpha [m_p z_o + f(1 - m_p) \cos \alpha]}{f(m_p \cos^2 \alpha + \sin^2 \alpha)}.$$

Furthermore, since it is rather difficult to invert Eq. (4.63) and Eq. (4.66) to obtain a formula for the lens tilt angle  $\alpha$  as a function of object tilt angle  $\beta$ , we use a computational technique to estimate  $\alpha$  starting from an initial estimate.

## Chapter 5

### SYNTHEZIZING EXTENDED DEPTH OF FIELD

*Technological innovation is combinatorial. New features of technology often arise by combining old things in new ways.*

—Andreas Wagner (*Arrival of the Fittest*)

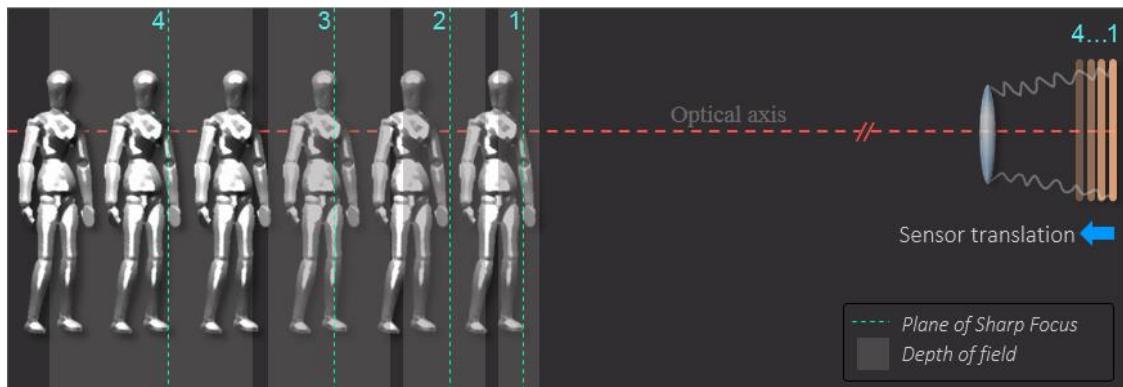
In the preceding chapters, we have learnt much about the geometric aspects of Scheimpflug imaging. How can we apply this knowledge to use Scheimpflug imaging for extending the depth of field (DOF) of iris acquisition systems? The obvious answer is that we orient the plane of sharp focus to pass through plane lying above the ground at the average height of the human eye, giving us infinite DOF along the longitudinal direction. However, because we would employ an open aperture to maximize spatial resolution attainable from the optics, the DOF will still be limited in the along the transverse plane. What could we do to increase the DOF along the transverse plane?

In this chapter, we will explore a specific computational imaging approach—focus stacking—and extend it to Scheimpflug imaging. We will see that by combining the central ideas of focus stacking with Scheimpflug imaging we can obtain a powerful solution that is very well suited—almost tailored—for extending the DOF of iris acquisition systems. We will explore the conditions that make the solution extremely simple and computationally efficient. We will also realize that the crux of the new technique lies in insights gained from our analysis of the geometric properties of the image in Scheimpflug cameras (Section 3.7)

## 5.1 Extending depth of field using frontoparallel focus stacking

The principle behind focus stacking (also known as z-stacking, multi-focus compositing, or focus blending) is simple. We capture a sequence of images while progressively focusing at axially increasing depths on the subject. The collection of images obtained is called a *focal stack* [82]. No single image in the stack contains the entire three-dimensional subject in focus. However, collectively, the stack includes all or most regions of the subject (or scene) in focus distributed amongst the images. An extended DOF image is created by selectively blending the sharp, in-focus areas into a single composite image.

There are several mechanisms for sweeping the focus across the object space. The two most methods are: (1) by translating either the sensor or the lens towards the other along the optical axis while keeping the focal length fixed as shown in [Figure 5.1](#), and (2) by changing the focal length. In the focus stacking method described above, the DOF regions are always perpendicular to the optical axis. We call this class of focus stacking technique as *frontoparallel focus stacking* (to distinguish it from the new method we are about to introduce.)



**Figure 5.1** Schematic of frontoparallel focus stacking. A sequence of images is captured while sweeping the focus axially across the subject (object space). In the above figure, the focus is changed by translating the sensor plane along the optical axis. The plane of sharp focus and DOF associated with each sensor position are also shown.

If the DOF extension is large, requiring a substantial change in sensor-to-lens plane distance or focal length, then the corresponding images of subjects in the stack will have significantly varying sizes owing to the difference in magnification. Additionally, if the camera is not held rigidly, slight movement in the camera will cause misalignment of corresponding subject images in the stack. Therefore, the images in the focal stack are often registered before generating the composite image.

The process of combining the images consists of detecting the in-focus regions in each image and blending the pixels across the images in the stack. At every pixel location in the composite image, a weighted average of all pixel intensities corresponding to the same position across the stack is often used. The filter weights are often determined based on an appropriate *focus measure* criterion. There are several focus measure criterions, but most algorithms use some form of local intensity contrast (image gradient) or local energy measure such as Laplacian of Gaussian (LoG) or Difference of Gaussian (DoG) filters [83–85]. Kumar and Ahuja’s paper on generative focus measure [86] is a good reference on subject of focus measure. As an example of the focus stacking method, [Figure 5.2](#) shows an extended depth of field composite image of a bug and three images (insets) from the corresponding focal stack demonstrated by Agarwala et al. in [87].



**Figure 5.2** Example of extended DOF in macro photography using frontoparallel focus stacking. The insets show three (out of thirteen) images from the focal stack obtained by varying the focus across the bug specimen. Note that in each of the inset images only a small region is in focus. The composite image obtained by selectively blending the in-focus regions from the images in the stack has the desired DOF atypical for macro photography. The focal stack images and the composite image were obtained from [87].

---

### 5.1.1 Advantages of focus stacking for extending depth of field

There are several advantages employing focus stacking to extend depth of field compared other methods:

1. Suppose we need  $n$  images, each exposed for  $t$  seconds, to generate the required DOF using focus stacking. Then, the total exposure time for the focus stacking based extended DOF image is  $T_{fs} = nt$  seconds. We could also obtain an equivalent DOF image with a single image capture (single-shot image) in a conventional camera by using a much smaller aperture (larger  $F/ \#$ ) value. However, as we have seen in [Sec. 1.4](#), using a smaller aperture

would result in lowering the exposure level (amount of light integrated by the sensor while the aperture is open). Therefore, to match the exposure level in the single-shot image to exposure level in the focus stacked image, we need to expose the single-shot image for a longer duration, say  $T_o$  seconds. Hasinoff and Kutulakos provided a proof in [83] which established that, in general,  $T_o > T_{fs}$ . In other words, the total exposure time required for focus stacking is less than the total exposure time required for single-shot image capture for the achieve equivalent exposure level and depth of field. Since larger exposure times may result in motion blur if subjects are in motion, we can argue that focus stacking is better suited than conventional single-shot imagery if subjects within an extended imaging volume is free to move.

2. As evidenced above, the typical tradeoffs between aperture, depth of field, and exposure time that constrain single-shot imagery does not apply to focus stack imagery [83]. The tradeoff between optical resolution versus depth of field for conventional optical imaging is also set loose. Whereas we need to use higher  $F/\#$  values to attain significant DOF in a single-shot image, a significantly lower  $F/\#$  value can be used for each segmented DOF region in focus stacking. Since the optical spatial resolution is inversely proportional to the  $F/\#$  (as shown in Eq. (1.1)), the composite image generated through focus stacking can attain significantly high optical resolution and depth of field at the same time (at the cost of increased focal stack size).
3. The quality of the captured image greatly depends on the exposure level. In the absence of thermal noise and sensor saturation, the pixel's signal-to-noise ratio increases with exposure level [83]. Since, we can use lower  $F/\#$  (bigger apertures) by chunking the required DOF, the SNR of (the in-focus regions in) each image in the focus stack and consequently the SNR of the composite image with extended DOF can attain high SNR.

Extending the DOF using frontoparallel focus stacking is not new. This imaging method appeared early in the field of microscopy [85,88] for imaging three-dimensional microscopic subjects wherein the DOF span is usually minuscule—typically between sub-millimeter and few millimeters owing to the high magnification. Adelson et al. [84] demonstrated the use of Laplacian pyramids to selectively blend in-focus regions to achieve significant DOF in macro photography. Since then several research efforts have spawned a multitude of techniques for fusing images from the stack obtained in frontoparallel focus stacking [87,89,90].

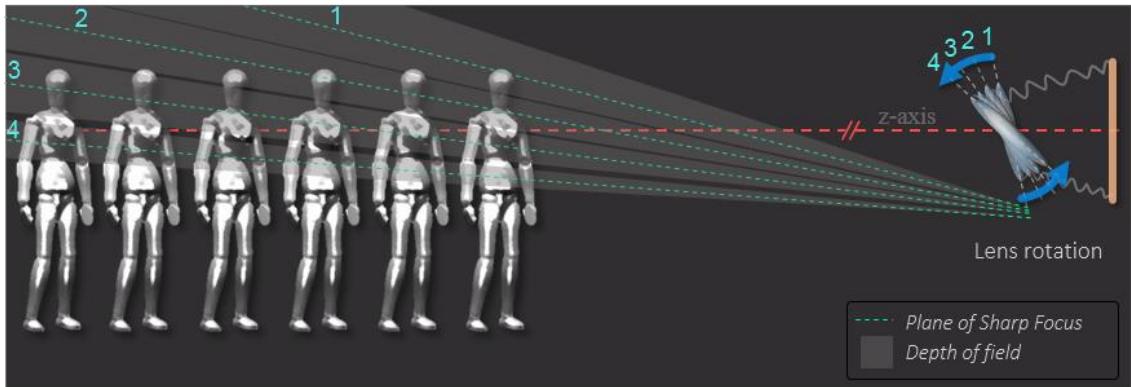
## 5.2 Extending depth of field using angular focus stacking

Given a subject with receding depth, Scheimpflug imaging allows us to orient the plane of sharp focus that maximizes the focus across the entire subject. However, the DOF is still limited to a finite region (approximately a wedge) surrounding the plane of sharp focus.

While the frontoparallel focus stacking method offers many attractive advantages, significant portions of the DOF in each image capture in the stack extends perpendicular to the optical axis and outside the field-of-view of the camera. Therefore, we can only achieve a suboptimal utilization of the set of DOF available in each capture.

A question naturally arises: is it possible to combine the mechanisms from Scheimpflug imaging and frontoparallel focus stacking for generating extended DOF images while bypassing the shortcomings of either method? The idea is illustrated in [Figure 5.3](#) in which we capture a sequence of images while the lens is continually rotated about a point along the optical axis. The continual rotation of the lens induces the plane of sharp focus (and the DOF surrounding it) to swing through the object space. Consequently, different parts of the scene come in and out of focus, governed by the Scheimpflug principle, across the images in the focal stack. Alternatively, we could rotate the sensor plane about its pivot to induce a rotation of the plane of sharp focus though the object space. Since the plane of focus sweeps across the object space angularly, we call this class of the focus stacking technique as *angular focus stacking*.

The advantage of angular focus stacking method over frontoparallel focus stacking is that a significant portion of the DOF for each image in the focal stack lie within the field-of-view of the camera. Therefore, we expect better utilization of the set of DOF available in each capture.



**Figure 5.3** Schematic of angular focus stacking. A stack of images is gathered while sweeping the focus angularly across the object space. The angular focus sweep is induced when the lens and/or the sensor planes are rotated about their independent pivots. If lens rotation is employed, as shown in this figure, then it is advantageous to pivot the lens at the center of the entrance pupil. This ensures there is no parallax between the images in the focal stack. Therefore, it permits us to align the images using analytic registration.

---

But how easy or difficult is it to do angular focus stacking? We have established in [Sec. 3.7](#) that rotating either the lens or sensor plane induces a warping of the image field. Particularly, rotation of the lens results in an additional shift of the image field. Therefore, before we can fuse the images in an angular focal stack we must register (align) the images in the stack.

Image registration is a critical step in angular focus stacking. Registration is the process of spatially aligning the images in the stack to a reference image by applying a mapping function—either known *a priori* from the model or estimated from the images. The degree of accuracy of image registration directly influences the quality of the synthesized image.

If only the sensor plane rotation is employed, then the images in the focus stack are related through a perspective projection. Therefore, we can use a perspective transformation to unwarp

each image to a chosen reference image. In the following discussion, we will concentrate on registering images that are captured while rotating the lens.

In general, rotation of the lens about a pivot along the optical axis results in a complex depth-dependent warping of the image field. In other words, different parts of the scene's image warp (distort and shift) by different amounts when the lens is rotated, depending on the corresponding part's depth in the object space. This is known as parallax. Although there are algorithms for registering images of the same scene exhibiting local variations, such registration methods are typically iterative in nature, and there are fundamental limits to the achievable registration accuracy [91], especially in the presence of noise and non-geometric distortions such as defocus blur. However, we have seen in [Sec. 3.7.3](#) that if we rotate the lens about the center of the entrance pupil, then the image field warping becomes independent of the scene depth. When the image field distortion is free of parallax effects, we can use a global mapping function to undistort the image. Furthermore, if the pupil magnification equals one, the images in the stack becomes pair-wise bilinear through a mapping  $H(\Delta\alpha):\hat{\mathbf{x}}_i \rightarrow \hat{\mathbf{x}}_j$ , where  $\Delta\alpha$  is the difference angle of the lens' orientation between  $\hat{\mathbf{x}}_i$  and  $\hat{\mathbf{x}}_j$ . This mapping is called the *inter-image homography* [92]. In [Sec. 5.2.1](#) we derive the inter-image homography for the case of rotating a unit pupil magnification lens about the entrance pupil starting from Eq. (3.27). The inter-image homography is used register the images in the focal stack analytically. Thus, the registration process becomes efficient (not requiring any iterative algorithm) and exact.

### **5.2.1 Inter-image homography for lens of unit pupil magnification, tilted about entrance pupil**

In this section, we derive the inter-image homography between images in the angular focal stack for the case when a lens, having unit pupil magnification, is tilted (rotation about the  $x$ -axis) about the center of the entrance pupil. The image plane is not tilted, and it is oriented perpendicular to the  $z$ -axis as shown in [Figure 5.3](#). We use the inter-image homography to register the images in the focal stack before fusing them to create a composite image having extended DOF.

The mapping between an object point  ${}^c\mathbf{x}$  (expressed in camera coordinates) and its image point  ${}^I\hat{\mathbf{x}}$  (in physical image coordinates) is as follows (Eq. (3.27)):

$${}^I\hat{\mathbf{x}} = R_i^T (\hat{d}_e \mathbf{r}_{\ell,3} - \mathbf{t}_i) + \frac{(\hat{\mathbf{n}}_i(3)\dot{z}_o - \hat{d}_e \hat{\mathbf{n}}_i^T \mathbf{r}_{\ell,3})}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T ({}^c\mathbf{x} - d_e \mathbf{r}_{\ell,3})} R_i^T R_\ell M_p R_\ell^T ({}^c\mathbf{x} - d_e \mathbf{r}_{\ell,3}) \quad (5.1)$$

Since the image plane is not rotated,  $R_i = I_{3 \times 3}$ , and  $\hat{\mathbf{n}}_i = 1$ . Further, since the lens is pivoted at the center of the entrance pupil,  $d_e = 0$ . If the directed distance from the entrance pupil to the exit pupil along the optical axis is  $d$ , then  $\hat{d}_e = d$ . Substituting,  $R_i = I$ ,  $\hat{\mathbf{n}}_i = 1$ ,  $d_e = 0$ , and  $\hat{d}_e = d$  in Eq. (5.1) we obtain:

$${}^I\hat{\mathbf{x}} = d \mathbf{r}_{\ell,3} - \mathbf{t}_i + \frac{\dot{z}_o - d R_\ell(3,3)}{\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T {}^c\mathbf{x}} R_\ell M_p R_\ell^T {}^c\mathbf{x} \quad (5.2)$$

Suppose the lens is tilted about the  $x$ -axis by an angle  $\alpha$ , where  $\alpha$  is numerically negative if the direction of rotation is from positive  $z$ -axis towards positive  $y$ -axis. Then, we can represent the orientation of the lens as:

$$R_\ell = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix}. \quad (5.3)$$

Further, we can write

$$R_\ell M_p R_\ell^T = \begin{bmatrix} {}^c_x \\ \{ {}^c_y(m_p \tan^2 \alpha + 1) + {}^c_z(1 - m_p) \tan \alpha \} \cos^2 \alpha \\ \{ {}^c_y(1 - m_p) \tan \alpha + {}^c_z(m_p + \tan^2 \alpha) \} \cos^2 \alpha \end{bmatrix}, \quad (5.4)$$

and

$$\hat{\mathbf{n}}_i^T R_\ell M_p R_\ell^T {}^c\mathbf{x} = \{ {}^c_y(1 - m_p) \tan \alpha + {}^c_z(m_p + \tan^2 \alpha) \} \cos^2 \alpha. \quad (5.5)$$

Substituting Eq. (5.5) into Eq. (5.2) yields:

$${}^I\dot{\mathbf{x}} = d \mathbf{r}_{\ell,3} - \mathbf{t}_i + \frac{\dot{z}_o - d \cos \alpha}{\mathcal{K}} R_\ell M_p R_\ell^T c_{\mathbf{x}} , \quad (5.6)$$

where,  $\mathcal{K} = \{c_y(1 - m_p) \tan \alpha + c_z(m_p + \tan^2 \alpha)\} \cos^2 \alpha$ . Note if  $\alpha = 0$ , then  $\mathcal{K} = c_z m_p$ .

Eq. (5.6) greatly simplifies when the pupil magnification is unity. Substituting  $m_p = 1$  in the above equation, we obtain:

$${}^I\dot{\mathbf{x}} = d \begin{bmatrix} 0 \\ -\sin \alpha \\ \cos \alpha - \dot{z}_o \end{bmatrix} + \frac{\dot{z}_o - d \cos \alpha}{c_z} \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} . \quad (5.7)$$

Expanding Eq. (5.7) to the constituent components, we obtain:

$$\begin{aligned} {}^I\dot{x} &= \frac{\dot{z}_o - d \cos \alpha}{c_z} c_x \\ {}^I\dot{y} &= \frac{\dot{z}_o - d \cos \alpha}{c_z} c_y - d \sin \alpha \\ {}^I\dot{z} &= 0 \end{aligned} \quad (5.8)$$

Note that the  $z$ -coordinate of the (image) points in the image reference frame  $\{I\}$  is zero. This is, of course, expected as the image coordinates are essentially two-dimensional. We will now represent Eq. (5.8) in matrix form as it will permit us to efficiently find an expression for the inter-image homography as a linear mapping between corresponding points in images from the focal stack. Furthermore, we will use *homogeneous coordinates* to represent the image points. That is, we add an extra dimension to the two-dimensional image coordinate by representing  $({}^I\dot{x}, {}^I\dot{y})$  as  $({}^I\dot{x}, {}^I\dot{y}, 1)$ . The use of homogeneous coordinates is a common practice in the Computer Vision literature as it allows us to represent affine transformations a linear mapping. The Eq. (5.8), in matrix form, with the image point represented in homogeneous coordinates, is as follows:

$$\underbrace{\begin{bmatrix} {}^I\dot{x} \\ {}^I\dot{y} \\ 1 \end{bmatrix}}_{{}^I\dot{\mathbf{x}}} = \frac{1}{c_z} \underbrace{\begin{bmatrix} (\dot{z}_o - d \cos \alpha) & 0 & 0 \\ 0 & (\dot{z}_o - d \cos \alpha) & -d \sin \alpha \\ 0 & 0 & 1 \end{bmatrix}}_{A(\alpha)} \underbrace{\begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix}}_{c_{\mathbf{x}}} . \quad (5.9)$$

Let the image in the angular focal stack corresponding to the lens in the nominal orientation ( $\alpha = 0^\circ$ ), be the reference image. Then, for an object point  ${}^c\mathbf{x}$ , we can obtain the following two mappings for two orientations of lens (one being the reference orientation):

$$\begin{aligned} {}^I\dot{\mathbf{x}}_\alpha &= A(\alpha) {}^c\mathbf{x} \\ \text{and} \\ {}^I\dot{\mathbf{x}}_0 &= A(0) {}^c\mathbf{x} \end{aligned} \quad (5.10)$$

Note that  ${}^I\dot{\mathbf{x}}_\alpha$  and  ${}^I\dot{\mathbf{x}}_0$  are the image points in frame  $\{I\}$  for the same object point  ${}^c\mathbf{x}$ . From Eq. (5.10) it follows that

$${}^I\dot{\mathbf{x}}_\alpha = \underbrace{A(\alpha) A(0)^{-1}}_{H(\alpha, 0)} {}^I\dot{\mathbf{x}}_0 . \quad (5.11)$$

The matrix  $H(\alpha, 0)$  in Eq. (5.11) is the inter-image homography. The inter-image homography  $H(\alpha, 0)$  is the pair-wise mapping between corresponding points in the two images of the same scene obtained under two orientations—the reference orientation ( $\alpha = 0^\circ$ ) and a finite tilt of the lens. Using Eqs. (5.9) - (5.11) we obtain the expression for  $H(\alpha, 0)$  as:

$$H(\alpha, 0) = \begin{bmatrix} \frac{(\dot{z}_o - d \cos \alpha)}{\dot{z}_o - d} & 0 & 0 \\ 0 & \frac{(\dot{z}_o - d \cos \alpha)}{\dot{z}_o - d} & -d \sin \alpha \\ 0 & 0 & 1 \end{bmatrix}, \quad (5.12)$$

where,

$\dot{z}_o$	Directed distance of the image plane's pivot from the camera frame, along the z-axis of the camera frame.
$d$	Directed distance of the exit pupil from the entrance pupil (lens pivot) along the optical axis.
$\alpha$	Tilt angle of the lens about the $x$ -axis.

### 5.2.2 Image registration using the inter-image homography

The images in the focal stack obtained under lens tilt (about the center of the entrance pupil) are related to the reference image (no lens rotation) as:

$${}^I\dot{\mathbf{x}}_\alpha = H(\alpha, 0) \ {}^I\dot{\mathbf{x}}_0 . \quad (5.13)$$

The  $3 \times 3$  matrix  $H(\alpha, 0)$  is a mapping of coordinates from the reference image to the image obtained under lens tilt. Therefore, for registering the image obtained under lens tilt to the reference image we apply the inverse matrix,  $H^{-1}(\alpha, 0)$ , to the coordinates of the image under lens tilt.

Note that we are not required to capture a reference image with  $\alpha = 0^\circ$ . But since we can register the images analytically, we align all images in the stack to the nominal lens orientation. Doing so ensures that the composite image is geometrically equivalent to a single-shot photograph of a given scene.

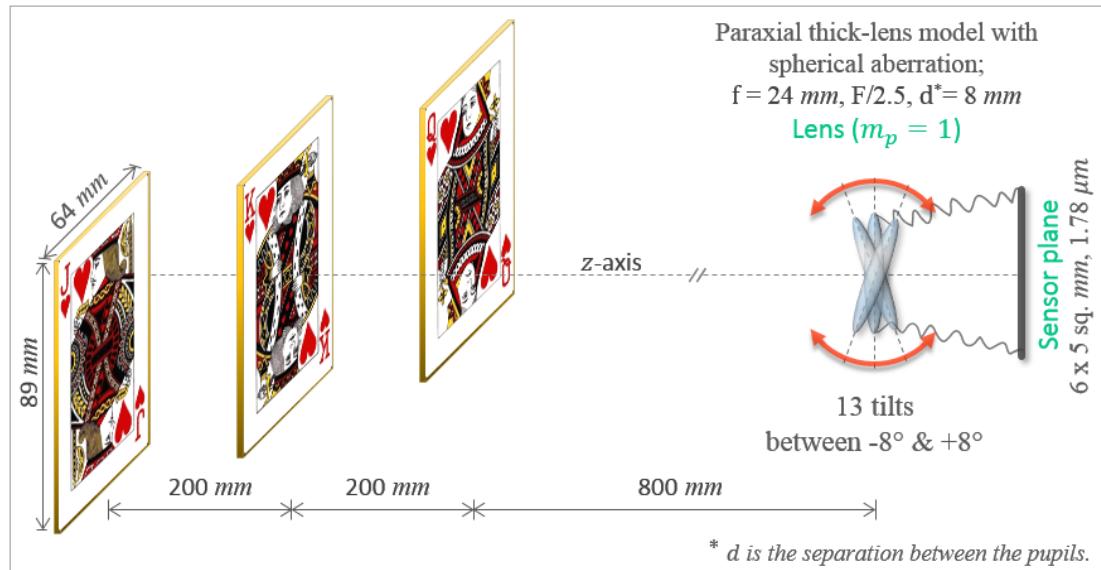
In the following section, we verify the above theory of synthesizing extended DOF images using angular focus stacking using a simulation setup in Zemax. Please note that our goal is not to present a new or best possible algorithm for detecting and fusing the in-focus regions from the images in the stack but rather to test the theory of angular focus stacking.

### 5.3 Simulation of extended DOF image synthesis using angular focus stacking

[Figure 5.4](#) shows a schematic of the image simulation setup in Zemax. We implemented a F/2.5 thick lens model using two paraxial surfaces of focal lengths  $f_1 = 40 \text{ mm}$  and  $f_2 = 30 \text{ mm}$  with separation  $s = 20 \text{ mm}$  between the two surfaces, resulting in an effective focal length  $f = 24 \text{ mm}$  ( $1/f = 1/f_1 + 1/f_2 - s/(f_1 f_2)$ ). A circular stop (diameter =  $7.14 \text{ mm}$ ) surface was placed behind the first paraxial surface at a distance  $a = 11.43 \text{ mm}$ , resulting in a pupil magnification  $m_p = 1$  ( $m_p = (f_2/f_1)((a - f_1)/(s - a - f_2))$ ). For tilting the object and lens independently, we set the object surface type as “Tilted,” and bracketed all surfaces associated with the lens within coordinate breaks.

The Image Simulation analysis tool in Zemax Sequential mode is powerful and offers an extensive set of tuning parameters. However, to produce an adequately representative simulation, the parameters must be chosen carefully based on the objective of the experiment. The most important parameters within the context of the current simulation are: (1) Field height of the source

bitmap, (2) Oversampling factor (if required), (3) Pupil sampling, (4) Image sampling, (5) Aberrations, (6) Reference, (7) Pixel size, and (8) X Pixels and Y Pixels. The image simulation in Zemax essentially consists of the three steps [93]: (a) The source bitmap image is convolved with a PSF grid (space variant and accounts for optical aberrations) generated in the object space whose fidelity depends on the set field height, oversampling factor and number of pixels; (b) The convolved image, in the object space, is transferred to the image space to account for geometric distortions and system magnification; and (c) The sampling effects of a discrete detector is simulated based on the set pixel size and detector size (inferred from pixel size and number of pixels). Since the paraxial surfaces are devoid of any aberrations, we inserted a Zernike Standard Phase surface at the location of the exit pupil to introduce slight spherical aberration. The small amount of spherical aberration also increased the spot size of the PSFs ensuring adequate pixels to represent each PSF. Additionally, we set sufficiently fine pupil sampling and image sampling (both 64 x 64) that influences how accurately the PSFs represent system aberrations.



**Figure 5.4** Schematic of simulation setup. The object space consists of three playing cards (89 mm x 64 mm) located at 0.8 m, 1 m, 1.2 m from the lens' vertex when the lens is not rotated (nominal orientation). We modelled a symmetric thick lens (pupil magnification  $m_p = 1$ ) of focal length  $f = 24 \text{ mm}$ ,  $F/\# = 2.5$  and pivoted it at the center of the entrance pupil. The focal stack consisted of thirteen images captured between  $\pm 8^\circ$  with uniform spacing in angles.

---

The scene for our simulation is three-dimensional consisting of three playing cards ( $64\text{ mm} \times 89\text{ mm}$ ) placed at  $800\text{ mm}$ ,  $1000\text{ mm}$  and  $1200\text{ mm}$  from the lens' vertex (before rotating the lens). However, the Image Simulation tool was not designed to simulate the imaging of three-dimensional scenes. Therefore, to emulate the effect of imaging a three-dimensional scene, we run the Image Simulation Tool in Zemax for each depth plane (three), with identical settings and integrated the outputs of each simulation into a single image. An obvious shortcoming of the simple integration process is that it fails to accurately simulate imaging portions of the scene where objects overlap in the image space. To avoid this problem, we spatially separated the three cards along the transverse direction (using appropriate fields setting in Zemax) such that their images (following blurring) does not overlap in the image plane (by picking 'Vertex' as the reference under detector settings). This limitation (and the workaround) does not, however, detract from the main purpose of the simulation—to test the feasibility of synthesizing an omnifocus image from a series of images captured under lens tilts.

To simulate imaging of a scene consisting of  $m$  depth planes for  $n$  orientations of the lens, we need to execute the Image Simulation tool  $m \times n$  times while setting the appropriate simulation parameters and integrating the  $m$  outputs for every orientation. We used PyZDDE [94] to automate the entire process of tilting the lens about the  $x$ -axis pivoted at the center of the entrance pupil to create a sequence of 13 images between  $\pm 8^\circ$ .

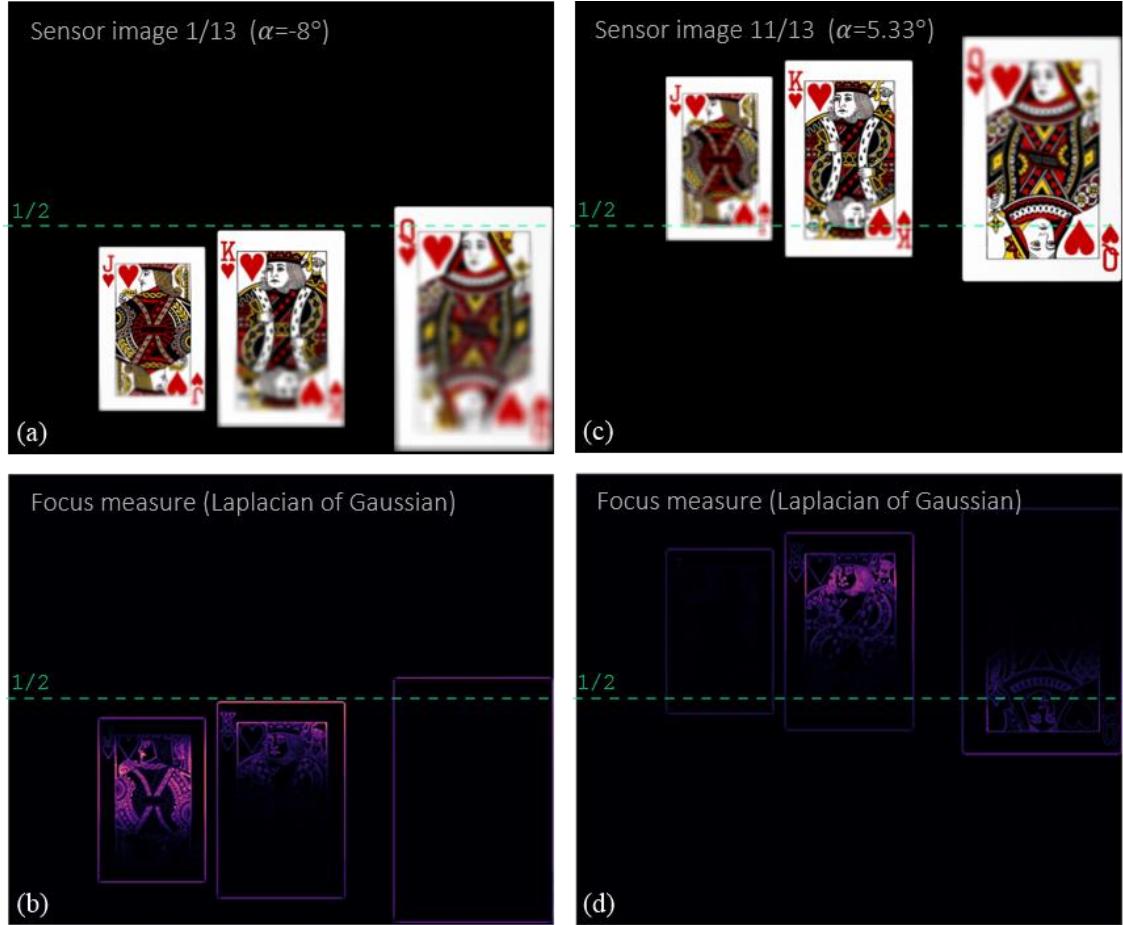
The integrated images of the three-dimensional scene for lens tilt angles  $\alpha = -8^\circ$  and  $\alpha = 5.55^\circ$  are shown in [Figure 5.5 \(a\)](#) and [Figure 5.5 \(c\)](#) respectively. Note the transverse shift of the image field in both images. Although not apparent in the figures, the individual images of the three cards in each integrated image in the sensor plane are vertically shifted and de-magnified by the same amount, as predicted by Eq. (5.12).

The in-focus regions in the integrated images, detected using a Laplacian of Gaussian (LoG) filter, are shown in [Figure 5.5 \(b\)](#) and [Figure 5.5 \(d\)](#). Note that no single plane is in complete (or uniform) focus, but parts of each plane that lie within the wedge shaped DOF surrounding the tilted plane of sharp focus form sharp regions in the sensor image.

The 13 integrated images were analytically registered using the inter-image homography matrix  $H(\alpha, 0)$  shown in Eq. (5.12). We used OpenCV's *warpPerspective* [95] (a geometric transformation function) that accepts the image under lens tilt, the homography matrix  $H(\alpha, 0)$  and returns the registered image. The registration process is pictorially depicted in [Figure 5.6](#).

Following registration, a composite image was created by blending the in-focus regions from the images. The process for selectively blending the in-focus regions, separately for each color plane  $k$ , to create the composite image  $I_k(m,n)$  can be represented as (borrowed from [85]):

$$I_k(m,n) = \sum_{p=1}^P I_{k,p}(m,n) S_p(m,n), \quad k = 1, 2, 3; \quad m = 1, \dots, M; \quad n = 1, \dots, N \quad (5.14)$$

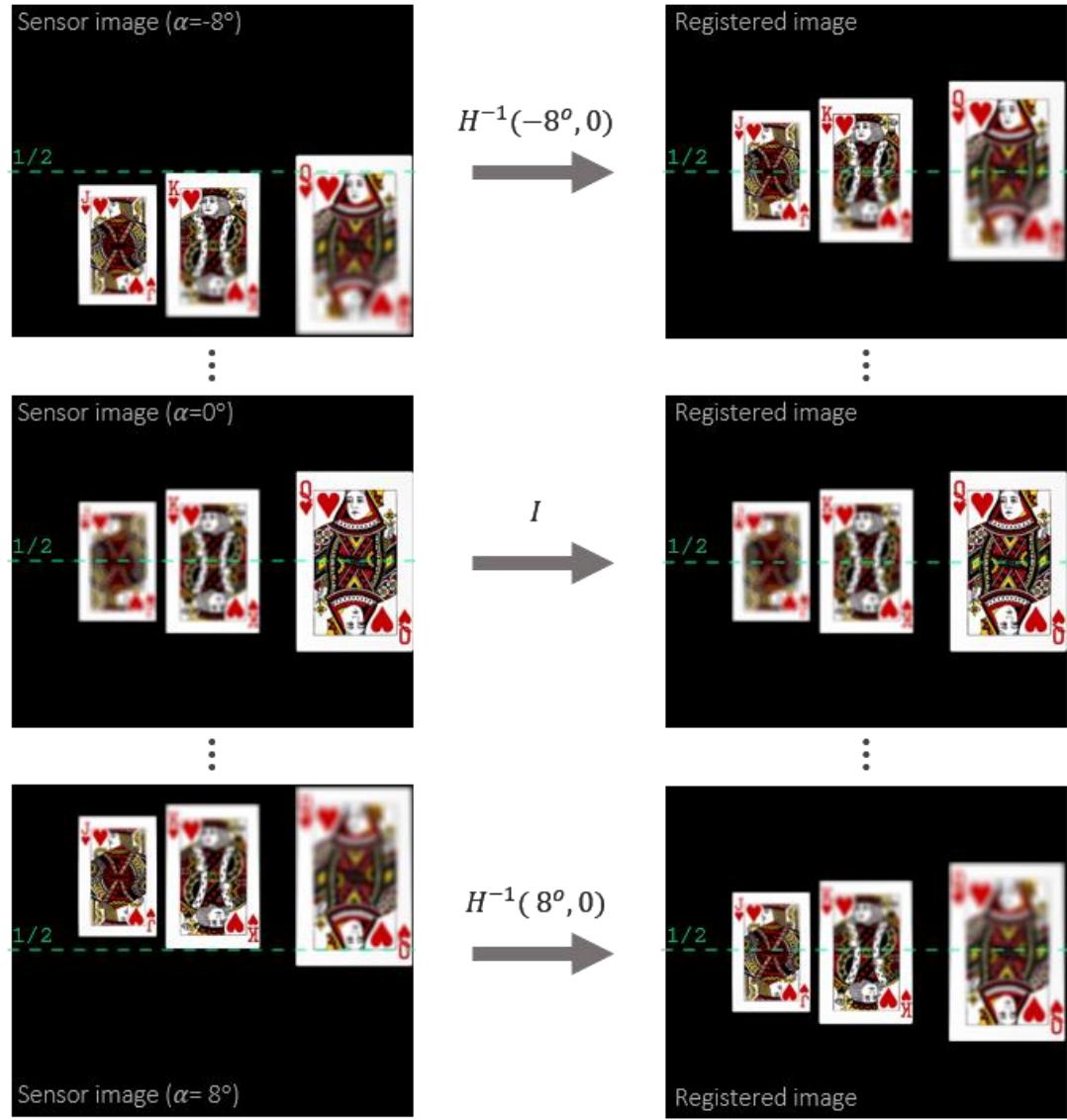


**Figure 5.5** Integrated sensor images (simulated) in the angular focal stack. (a) Sensor image for lens tilted by angle  $\alpha = -8^\circ$  about the  $x$ -axis. (b) The focus measure of the sensor image in (a) showing the in-focus regions. The in-focus regions were detected using a Laplacian of Gaussian (LoG) filter. (c) Sensor image for lens tilted by angle  $\alpha = 5.33^\circ$  about the  $x$ -axis. (d) The focus measure of the sensor image in (c). The overlaid green dashed lines indicate the middle of sensor in the vertical direction. It gives a visual reference for comparing the vertical shift in each sensor image induced by lens rotation.

---

where,

$I_k(m, n)$	The $k^{th}$ color channel (red, green, or blue) of the composite image. $k = 1, 2, 3$ .
$I_{k,p}(m, n)$	The $k^{th}$ color channel of the $p^{th}$ image in the focal stack of size $P$ . Each image in the focal stack is an array of size $M \times N \times 3$ . Where $M$ is the number of rows, and $N$ is the number of columns.
$S_p(m, n)$	The image sampling function. The form of the image sampling function depends on the type of algorithm that is used to assign weights to the images in the focal stack for each pixel location $(m, n)$ .



**Figure 5.6** Analytic registration of images in the focal stack. The images on the left column shows the integrated sensor data for lens tilts  $\alpha = -8^\circ, 0^\circ, 8^\circ$ . The rotation of the lens about the entrance pupil induces a transverse shift and uniform scaling of the image field. The images on the right column shows the analytically registered images, aligned with respect to the reference ( $\alpha = 0^\circ$ ) image, using Eq. (5.12).

---

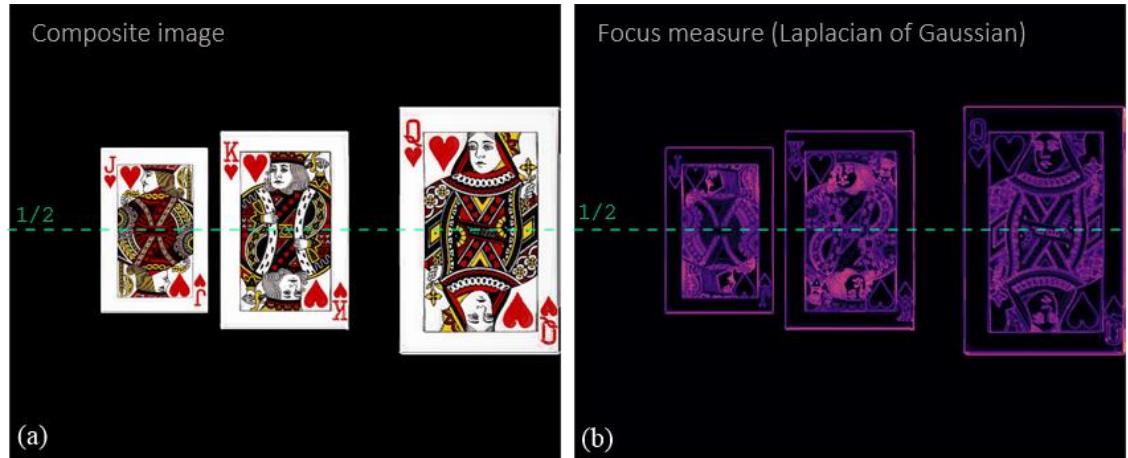
There are several sophisticated techniques for fusing sharp regions from registered images in a focal stack [87,89]. However, our goal is to demonstrate the feasibility of angular focus stacking as an alternative approach for omnifocus imagery. Moreover, using a simple image sampling function can reveal any artifacts of the analytic registration that is critical to our approach.

Therefore, we have opted to use the following simple image sampling function (as described by Streibl in [85]):

$$S_p(m,n) = \delta_{pp_o} = \begin{cases} 1, & p = p_o(m,n) \\ 0, & p \neq p_o(m,n) \end{cases} \quad (5.15)$$

where,  $p_o(m,n)$  represents the image number in the focal stack that maximizes a focus measure function at the pixel location  $(m,n)$ . Similar to the image sampling function, there are several choices for available for the focus measure function. As discussed previously, we have used the LoG filter as an indicator of focus in our algorithm.

[Figure 5.7 \(a\)](#) shows the composite image in which the complete scene consisting of the three planes is in focus. [Figure 5.7 \(b\)](#) shows the degree of focus on the three planes in the composite image, measured using the LoG filter. The figure demonstrates that all three planes at different distances from the lens is in focus in the computationally generated image.



**Figure 5.7** Result of the angular focus stacking simulation in Zemax. (a) The composite image obtained by selectively combining the in-focus regions within the images in the focal stack. (b) Focus measure of the composite image proves that entire three-dimensional scene is in focus in the synthesized image.

#### **5.4 Advantages of angular focus stacking for extending the DOF of iris acquisition systems**

Angular focus stacking shares all the advantages with frontoparallel focus stacking enumerated in [Sec. 5.1.1](#). Additionally, angular focus stacking is especially conducive for extending the imaging volume of iris acquisition systems. For iris recognition, we are often interested in capturing high-resolution images within the capture volume, approximately the shape of a rectangular parallelepiped. We can define the elevation of the parallelepiped as the average height of the human eye from the ground. The length of the parallelepiped is specified by the depth (range of distances from the camera) over which we want to perform iris acquisition. The width is usually bounded by the field-of-view of the camera in the direction orthogonal to the depth and height. We could define the height of the parallelepiped in several ways: for example, we could set the height as a function of the standard deviation of human height. Once the parallelepiped has been defined, we can orient the plane of sharp focus along the length of the parallelepiped using Scheimpflug imaging (using the set of equations derived in Chapter 4). Further, we can employ angular focus stacking to increase the DOF to fill the height of the parallelepiped. For a fixed focal length and aperture, the number of DOF segments required depends on the prescribed optical resolution. While the plane of sharp focus and the DOF can extend to infinity along the depth (for an adequate amount of lens tilt), the far extent of the parallelepiped along its length limited by the system magnification. The system magnification determines the total number of pixels that span across the iris image in the sensor plane.

#### **5.5 Summary**

In this chapter, we invented a computational technique for synthesizing an extended DOF image by selectively blending the in-focus regions from a stack of images captured while rotating the lens about the entrance pupil. We arrived at this solution, which we called *angular focus stacking*, by combining the ideas of conventional focus stacking and Scheimpflug imaging. Starting from our geometric model of Scheimpflug imaging, we derived a closed form expression for the inter-image

homography which we use to analytically register the images in the stack. The ability to register the images analytically is crucial not only because of the absence of optical blur in the images but also because it makes registration efficient and exact. Moreover, we saw that if use a lens with unit pupil magnification, then the inter-image homography is simple combination of uniform scaling and translation (in the image coordinate frame). Further, we saw that angular focus stacking is especially suitable for extending the depth of field of iris acquisition systems.

## Chapter 6

### DISCUSSION

*When something happens that you don't understand [...] it is worth spending time to figure out what's the weirdness? It's out of problems that you can't understand that you make progress.*

—Cliff Stoll (Good Science, Tech Icons)

#### 6.1 Summary of the work

Iris recognition has the potential to fulfill the technological need for providing security and surveillance. However, due of the fundamental limitation of depth of field imposed by the physics of light, contemporary iris acquisition systems fail to capture high-quality iris images outside a small capture volume. This shortcoming has significantly impeded the use of iris recognition in large scale, unconstrained environments. The goal of this thesis was to investigate the depth of field problem pertinent to iris recognition and propose a scalable and computationally efficient solution that can significantly improve the capture volume.

Following a comprehensive survey of the state-of-the-art techniques for improving the capture volume of iris acquisition systems, we found that there are broadly two main types of solutions to increase the capture volume significantly. The first type of solutions combines multiple field of views (FOVs) either in time or space to expand the total field of view of the iris acquisition system. For example, multiple cameras (with a variety of focal lengths) may be used to stagger the individual FOVs in space, or a single camera with a zoom lens and focusing unit may be mounted on a pan-tilt stage to cover the desired capture volume over time. These solutions tend to be

complex in operation and are usually expensive. The second type of solutions employs computational imaging techniques, such as wavefront coding, to make the impulse response (point spread function) greatly invariant to defocus so that a suitable inverse filter can be used to recover fine details (high spatial frequency) information from the scene across a considerable depth. However, these systems are usually plagued by noise exactly at the high spatial frequencies and tend to have high computational complexity.

In this work, we investigated Scheimpflug along with computational imaging as a possible approach for extending the depth of field for iris acquisition. Scheimpflug imaging is an age-old technique used by landscape photographers to capture sharp images of scenes that have very close foreground subjects and receding background subjects in a single shot. Although Scheimpflug imaging doesn't increase the depth of field of the imager, changing the orientation of the plane of sharp focus (and the depth of field) gives the impression of extended depth of field.

We developed a pair of imaging models in this work that enabled us to study and predict the geometric properties of the image formed in imaging systems in which the lens and the sensor are free to rotate about their independent pivots. The new models are no more or no less accurate than existing thick lens models of Scheimpflug imaging that abstract the lens using the principal planes. Nevertheless, our models incorporate a different set of parameters—the pupil locations and pupil magnification—which allows the models to predict the image field distortion upon lens (or sensor) rotation more efficiently than existing models. We verified the accuracy of the models by comparing the image coordinates numerically computed using our models against the image coordinates obtained through ray tracing in Zemax for randomly generated optical systems (described by the pupil parameters). Further, we showed, in [Sec. 4.2](#), that our general model reduces to simpler relationships for known and specific configurations of Scheimpflug imaging. We also presented a fast, iterative algorithm to determine the required the lens tilt angle and sensor plane distance for focusing on a given tilted object surface.

Furthermore, a geometric analysis of the image using our model lead to the discovery of the set of conditions required for computationally synthesizing an extended depth of field image from images collected while rotating a lens about the center of the entrance pupil. We call this new computational imaging technique as *angular focus stacking*. Although we invented the angular focus stacking method by combining ideas from Scheimpflug imaging and the conventional focus stacking, the central concept of the method originated from the insights gleaned from the new geometric model. Theoretically, the extension of depth of field using the angular focus stacking method is unbounded; however, the number of images in the focal stack increases with the extents of the desired depth of field. In [Sec. 5.3](#) we demonstrated the angular focus stacking method using a simulation created with Zemax and PyZDDE.

The angular focus stacking method shares all the advantages of conventional camera focus stacking (or frontoparallel focus stacking), mainly that it is not restricted by the traditional tradeoffs between optical resolution, exposure level, and depth of field. As shown in [Sec. 5.2](#), the stack of images obtained by rotating the lens about the center of the entrance pupil can be analytically registered using the inter-image homography derived from our model. The capability to do analytic registration is a huge advantage because registration is exact and efficient even in the presence of noise. The angular focus stacking technique is particularly suitable for significantly increasing the capture volume of iris acquisition systems as argued in [Sec. 5.4](#); yet it is simple, easily scalable and computationally efficient.

## 6.2 Conclusion

We have invented a simple method that can significantly (in theory, unboundedly) extend the depth of field of iris acquisition systems.

## 6.3 Limitations

In this work, we developed geometric models of Scheimpflug imaging that helped us to device a simple, yet efficient computational technique for extending the depth of field of iris acquisition

systems. To make the modeling tractable, we assumed the optics to behave rectilinearly (having no radial distortions), and free of optical aberrations. Further, we did not incorporate the effects of diffraction on the bounds of the depth of field in Scheimpflug imaging. However, the absence of diffraction analysis does not have direct bearing on the goals and outcome of the study (i.e. to investigate the feasibility of using Scheimpflug imaging for extending the capture volume of iris acquisition systems) if we use fairly well-corrected optics in the field because: (1) the subject distances are much greater than the aperture of the lens implying that the diffraction can be neglected for estimating the depth of field for all practical purposes, (2) whereas the effects of diffraction becomes more pronounced at smaller apertures, we expect to use the lens is open aperture (or an optimum aperture) such that the effects of diffraction can be neglected.

The strategy currently employed for gathering images in the angular focal stack is by rotating the lens by uniform angular steps. While this strategy works if we would like increase the depth of field only by a finite amount about the tilted plane of sharp focus, the strategy may be sub-optimum for the goal is to improve the depth of field over a large volume, for example for synthesizing an omnifocus (all-in-focus) image of general three-dimensional scene. Therefore, an optimization strategy must be used to find the minimum number of lens rotations and the set of lens rotation angles constrained upon the imaging volume, aperture setting and desired spatial resolution.

#### **6.4 Direction of future research**

Range estimation or gathering three-dimensional information about the scene, such as the capability to generate a depth map is an active area of research. Since the center of the entrance pupil is the center of perspective projection, we have seen that images captured while rotating the lens about a point away from the center of the entrance pupil exhibit parallax due the change in viewpoint. In other words, we observe a depth dependent deformation of image field. Can we, then, use the geometric model developed in Chapter 3 to passively (without structured illumination) estimate the depth of scene-elements from the knowledge of the deformation? In depth from stereo geometry,

the accuracy of depth estimation is directly proportional to the baseline (i.e. the separation between the center of the entrance pupils of the two cameras).

If we pivot the lens about the center of the entrance pupil, the viewpoint doesn't change when we rotate the lens. However, the plane of sharp focus sweeps through the scene per the Scheimpflug principle. Is it possible to generate a depth map using the information about the amount of defocus in the images in the angular focal stack akin to what is done in depth from focus/defocus?

Although we have left out diffraction analysis for this work (as discussed in [Sec. 6.3](#)), a diffraction analysis will certainly provide a richer description of Scheimpflug imaging. Such, a rich description can expand the areas of application of Scheimpflug imaging. Moreover, the geometric model developed in this work can be used as a starting point because the model already incorporates the pupil parameters (location of the entrance and exit pupils and their size) that are essential for diffraction based analysis of optical systems.

Furthermore, there are two important consequences of the lens not being parallel to the sensor in Scheimpflug configuration. The first consequence is that the optical system becomes space variant, even in the absence of optical aberrations and geometric distortions, because of the variance in the distance between the exit pupil and the focused image points across the sensor plane. In other words, the diffraction limited spot size varies as a function of field across the sensor. The second consequence is that the shape of the point spread function is altered due to diffraction on the titled image surface (with respect to the exit pupil). For example, if we have a circular aperture, the Airy disk pattern observed on the image plane when it is parallel to the lens plane becomes elliptical in Scheimpflug imaging configuration. Based on diffraction analysis, is it possible to find a complex pupil mask that optically makes the system space invariant?

The ability to freely rotate (and translate) the lens and sensor planes about independent pivots provides much greater degrees of freedom for imaging than just the ability to focus on tilted object

surfaces. For instance, tilting the sensor while keeping the lens fixed produces images (of the same subject) that have slightly different perspective projection without any parallax (see [Sec. 3.7.1](#)).

How can these images be combined to generate a digitally super-resolved image?

We have investigated a method for synthesizing an omnifocus (all-in-focus) image from images captured while tilting the lens. However, angular focus stacking, in principle, should work even with images captured while rotating the sensor. Even for this case, an analytic registration method can be used to align the images in the stack.

## APPENDIX A

### Appendix A.1 Transfer of chief ray's direction cosine for arbitrary orientation of the optical axis

In [Sec. 3.4](#) we derived the expression for the transfer of the chief ray's direction cosine from the entrance pupil to the exit pupil for a specific problem in which the optical axis was coincident with the  $z$ -axis of  $\{C\}$ . Furthermore, we *inferred* the general expression for the transfer relation—in which the optical axis is free to swivel about the origin of  $\{C\}$ —from the expression obtained for the specific problem. Here we apply the method of induction to formally *derive* the general expression.

Eq. (3.10) accurately represents the *transfer* for the specific problem; however, we will cast the expression in a slightly different form whose *raison d'être* is to enable generalization—through direct application of the result. Specifically, we express the output chief ray as a linear combination of the input chief ray and the optical axis; this is possible because the two rays and the optical axis span the same (meridional) plane. Let  $\hat{\mathbf{c}}_z$ , the standard basis vector along  $z$ -axis of  $\{C\}$ , represent the optical axis since the optical axis is coincident with the  $z$ -axis. Then,

$$\hat{\mathbf{l}} = w_1 \mathbf{l} + w_2 \hat{\mathbf{c}}_z , \quad (A1.1)$$

where  $w_1$  and  $w_2$  are weights, and  $\hat{\mathbf{c}}_z = [0, 0, 1]^T$ .

Rewriting the above equation as:

$$\begin{bmatrix} \hat{l} \\ \hat{m} \\ \hat{n} \end{bmatrix} = w_1 \begin{bmatrix} l \\ m \\ n \end{bmatrix} + w_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} , \quad (A1.2)$$

we can readily obtain the weight  $w_1$  by comparing equations Eq. (3.9) and Eq. (A1.2) as:

$$w_1 = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n^2}} . \quad (A1.3)$$

Substituting  $w_1$  into  $\dot{n} = w_1 n + w_2$  and comparing with Eq. (3.8) yields  $w_2$  as:

$$w_2 = \pm \frac{(m_p - 1)n}{\sqrt{1 + (m_p^2 - 1)n^2}} . \quad (A1.4)$$

We are now ready to apply the result of the specific problem to the general problem. [Figure 3.5](#) shows the schematic of the general problem—the optical axis pivots about the origin of  $\{\mathcal{C}\}$ . Let us describe the general orientation of the optical axis by the action of the rotation matrix  ${}^cR_\ell \in \mathbb{R}^{3 \times 3}$  on  $\hat{\mathbf{c}}_z$ . The matrix  ${}^cR_\ell$  may be a composition of two or more matrices that denotes a sequence of rotations about the  $x$ -axis and  $y$ -axis. Then,  $\hat{\mathbf{o}}$ , the unit vector representing the new orientation of the optical axis, is obtained as the transformation of  $\hat{\mathbf{c}}_z$  by the rotation matrix:  $\hat{\mathbf{o}} = {}^cR_\ell \hat{\mathbf{c}}_z$  or  $\hat{\mathbf{c}}_z = ({}^cR_\ell)^T \hat{\mathbf{o}}$ .

Since the output direction cosine  $\dot{\mathbf{l}}$ , the input direction cosine  $\mathbf{l}$ , and the optical axis  $\hat{\mathbf{o}}$  lie on the same plane we can write  $\dot{\mathbf{l}}$  as the linear combination of  $\mathbf{l}$  and  $\hat{\mathbf{o}}$ :

$$\dot{\mathbf{l}} = w_1 \mathbf{l} + w_2 \hat{\mathbf{o}} . \quad (A1.5)$$

Note that the input direction cosine  $\mathbf{l}$  in Eq. (A1.5) (following the rotation of the optical axis) is, in general, different from the corresponding  $\mathbf{l}$  in Eq. (A1.1) even for the same object-point  $\mathbf{x}$ . This difference is due to the displacement of entrance pupil ( $E$ ) following the rotation of the optical axis; in fact, the designation of a ray as the chief ray (from  $\mathbf{x}$  to  $E$ ) alters as we displace the entrance pupil. Multiplying Eq. (A1.5) by  $({}^cR_\ell)^T$ , we obtain:

$$({}^cR_\ell)^T \dot{\mathbf{l}} = w_1 ({}^cR_\ell)^T \mathbf{l} + w_2 ({}^cR_\ell)^T \hat{\mathbf{o}} . \quad (A1.6)$$

Letting  $({}^cR_\ell)^T \dot{\mathbf{l}} = \dot{\mathbf{l}}_R$  and  $({}^cR_\ell)^T \mathbf{l} = \mathbf{l}_R$ , yields

$$\dot{\mathbf{l}}_R = w_1 \mathbf{l}_R + w_2 \hat{\mathbf{c}}_z . \quad (A1.7)$$

Comparing Eqs. (A1.1) and (A1.7), we obtain the expressions for the weights  $w_1$  and  $w_2$  as:

$$w_1 = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} ,$$

and

$$w_2 = \pm \frac{(m_p - 1)n_R}{\sqrt{1 + (m_p^2 - 1)n_R^2}} .$$
(A1.8)

Where  $n_R$  represents the projection of the direction cosine vector,  $\mathbf{l}$ , on the rotated optical axis. If we write the matrix  ${}^cR_\ell = [{}^c\mathbf{r}_{\ell,1} \quad {}^c\mathbf{r}_{\ell,2} \quad {}^c\mathbf{r}_{\ell,3}]$  where  ${}^c\mathbf{r}_{\ell,i}$  for  $i = 1, 2, 3$  are the columns of  ${}^cR_\ell$ , then:

$$({}^cR_\ell)^T = [({}^c\mathbf{r}_{\ell,1})^T \quad ({}^c\mathbf{r}_{\ell,2})^T \quad ({}^c\mathbf{r}_{\ell,3})^T]^T$$

and

$$\begin{aligned} \mathbf{l}_R &= ({}^cR_\ell)^T \mathbf{l} \\ &= [({}^c\mathbf{r}_{\ell,1})^T \quad ({}^c\mathbf{r}_{\ell,2})^T \quad ({}^c\mathbf{r}_{\ell,3})^T]^T \mathbf{l} \\ &= [({}^c\mathbf{r}_{\ell,1})^T \mathbf{l} \quad ({}^c\mathbf{r}_{\ell,2})^T \mathbf{l} \quad ({}^c\mathbf{r}_{\ell,3})^T \mathbf{l}]^T \end{aligned}$$

Therefore,  $n_R = ({}^c\mathbf{r}_{\ell,3})^T \mathbf{l}$  since  $n_R$  is the third element of  $\mathbf{l}_R$ .

Rewriting Eq. (A1.7) as:

$$\begin{aligned} \mathbf{l}_R &= \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} \mathbf{l}_R \pm \frac{(m_p - 1)n_R}{\sqrt{1 + (m_p^2 - 1)n_R^2}} \hat{\mathbf{e}}_z \\ &= \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} \left( \begin{bmatrix} l_R \\ m_R \\ n_R \end{bmatrix} \pm (m_p - 1)n_R \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \end{aligned}$$

$$\hat{\mathbf{l}}_R = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & m_p \end{bmatrix}}_{M_p} \begin{bmatrix} l_R \\ m_R \\ n_R \end{bmatrix}$$

which can be compactly written as:

$$\hat{\mathbf{l}}_R = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} M_p \mathbf{l}_R . \quad (A1.9)$$

Finally, substituting  $({}^c R_\ell)^T \hat{\mathbf{l}} = \hat{\mathbf{l}}_R$  and  $({}^c R_\ell)^T \mathbf{l} = \mathbf{l}_R$  yields the general expression for the direction cosines of the chief ray in the image space in terms of the pupil magnification and direction cosines in the object space as:

$$\hat{\mathbf{l}} = \pm \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} {}^c R_\ell M_p ({}^c R_\ell)^T \mathbf{l} . \quad (A1.10)$$

where  $n_R = ({}^c \mathbf{r}_{\ell,3})^T \mathbf{l}$

## Appendix A.2 The direction cosine, originating from exit pupil, has unit $\ell^2$ -Norm

Claim: The direction cosine  $\hat{\mathbf{l}}$  in the image space, obtained by the linear transformation of the direction cosine  $\mathbf{l}$  in the object space, has unit  $\ell^2$ -Norm, and  $\{1 + (m_p^2 - 1)n_R^2\}^{-1/2}$  is the normalization term.

*Proof.*

The expression for the direction cosine in the image space is

$$\hat{\mathbf{l}} = \frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}} {}^cR_\ell M_p ({}^cR_\ell)^T \mathbf{l} . \quad (A2.1)$$

where  $n_R = ({}^c\mathbf{r}_{\ell,3})^T \mathbf{l}$ ,  ${}^c\mathbf{r}_{\ell,i}$  is the  $i^{th}$  column of the rotation matrix  ${}^cR_\ell \in \mathbb{R}^{3 \times 3}$  applied to the optical axis,  $M_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & m_p \end{bmatrix}$ , and  $m_p \in \mathbb{R}^+$  is the pupil magnification.

Our objective is to prove  $\|\hat{\mathbf{l}}\|^2 = 1$ . For the convenience of notation within the proof, let

$$A = {}^cR_\ell M_p ({}^cR_\ell)^T = Q \Lambda Q^T , \quad (A2.2)$$

where  $Q = {}^cR_\ell$ , and  $\Lambda = M_p$ , the diagonal matrix with non-negative real values.

Also, we represent the columns of  $Q$  as  $\mathbf{q}_i = {}^c\mathbf{r}_{\ell,i}$  for  $i = 1, 2, 3$ . Then,  $n_R = ({}^c\mathbf{r}_{\ell,3})^T \mathbf{l} = \mathbf{q}_3^T \mathbf{l}$ , and

$$1 + (m_p^2 - 1)n_R^2 = 1 + (m_p^2 - 1)(\mathbf{q}_3^T \mathbf{l})^2 . \quad (A2.3)$$

Since  $Q$  is a rotation matrix, it is orthonormal (the columns of  $Q$ , having unit length, are orthogonal to each other). Therefore,  $Q^T Q = Q Q^T = I$ . Then,

$$\|\hat{\mathbf{l}}\|^2 = \frac{1}{1 + (m_p^2 - 1)(\mathbf{q}_3^T \mathbf{l})^2} \|A \mathbf{l}\|^2 , \quad (A2.4)$$

where

$$\begin{aligned}
\|A\mathbf{l}\|^2 &= (A\mathbf{l})^T(A\mathbf{l}) \\
&= (Q\Lambda Q^T\mathbf{l})^T(Q\Lambda Q^T\mathbf{l}) \\
&= \mathbf{l}^T Q\Lambda^T Q^T Q\Lambda Q^T \mathbf{l} \\
&= \mathbf{l}^T Q\Lambda^2 Q^T \mathbf{l} \\
&= \mathbf{s}\Lambda^2 \mathbf{s}^T
\end{aligned}$$

and  $\mathbf{s} = Q^T\mathbf{l} = [\mathbf{q}_1^T \mathbf{l} \quad \mathbf{q}_2^T \mathbf{l} \quad \mathbf{q}_3^T \mathbf{l}]^T \in \mathbb{R}^{3 \times 1}$  and  $\mathbf{s}(i) = \mathbf{q}_i^T \mathbf{l} \in \mathbb{R}$ .

Since  $\Lambda^2$  is a diagonal matrix, we can rewrite  $\|A\mathbf{l}\|^2$  as

$$\begin{aligned}
\|A\mathbf{l}\|^2 &= \sum_{i=1}^3 |\mathbf{s}(i)|^2 \Lambda^2(i,i) \\
&= (\mathbf{q}_1^T \mathbf{l})^2 \cdot 1 + (\mathbf{q}_2^T \mathbf{l})^2 \cdot 1 + (\mathbf{q}_3^T \mathbf{l})^2 \cdot m_p^2 .
\end{aligned} \tag{A2.5}$$

$$\text{Now, } \mathbf{s}^2 = \mathbf{s}^T \mathbf{s} = (Q^T \mathbf{l})^T (Q^T \mathbf{l}) = \mathbf{l}^T Q^T Q \mathbf{l} = \mathbf{l}^T \mathbf{l} = 1 \quad (\because \|\mathbf{l}\|^2 = 1, \text{ by definition}) . \tag{A2.6}$$

Also,

$$\begin{aligned}
\mathbf{s}^2 &= [\mathbf{q}_1^T \mathbf{l} \quad \mathbf{q}_2^T \mathbf{l} \quad \mathbf{q}_3^T \mathbf{l}] \begin{bmatrix} \mathbf{q}_1^T \mathbf{l} \\ \mathbf{q}_2^T \mathbf{l} \\ \mathbf{q}_3^T \mathbf{l} \end{bmatrix} \\
&= (\mathbf{q}_1^T \mathbf{l})^2 + (\mathbf{q}_2^T \mathbf{l})^2 + (\mathbf{q}_3^T \mathbf{l})^2 \\
&= 1 \quad (\because \mathbf{s}^2 = 1, \text{ from above}) .
\end{aligned} \tag{A2.7}$$

Substituting  $(\mathbf{q}_1^T \mathbf{l})^2 + (\mathbf{q}_2^T \mathbf{l})^2 = 1 - (\mathbf{q}_3^T \mathbf{l})^2$  in Eq. (A2.5) we obtain

$$\|A\mathbf{l}\|^2 = 1 - (\mathbf{q}_3^T \mathbf{l})^2 + (\mathbf{q}_3^T \mathbf{l})^2 \cdot m_p^2 = 1 + (m_p^2 - 1)(\mathbf{q}_3^T \mathbf{l})^2 . \tag{A2.8}$$

Further, substituting  $\|A\mathbf{l}\|^2$  into Eq. (A2.4) we obtain

$$\|\mathbf{l}\|^2 = \frac{1}{1 + (m_p^2 - 1)(\mathbf{q}_3^T \mathbf{l})^2} \|A\mathbf{l}\|^2 = 1 . \tag{A2.9}$$

It follows that the scalar quantity  $\frac{1}{\sqrt{1 + (m_p^2 - 1)n_R^2}}$  is the normalization term.  $\square$

## APPENDIX B

### **Appendix B.1 Derivation of Gaussian imaging equation with pupil magnification**

The familiar Gaussian imaging equation,  $-1/u + 1/\bar{u} = 1/f$ , relates the object and image plane distances with the focal length  $f$ . In this formula,  $u$  is the *directed distance* (numerically negative as per our sign convention) between the object plane (perpendicular to the optical axis) and the principal plane ( $H$ ) in the object space,  $\bar{u}$  is the directed distance (numerically positive for *real* images) between the in-focus image plane and the principal plane ( $\bar{H}$ ) in the image space. The distances  $u$  and  $\bar{u}$  are measured along the optical axis.

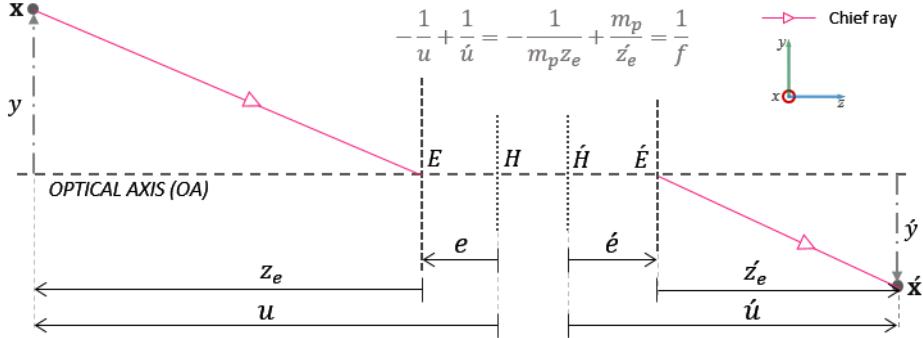
If the distances of the object and image planes are specified with respect to the entrance ( $E$ ) and exit ( $\bar{E}$ ) pupil centers instead of the Principal planes, then the Gaussian lens formula needs to be modified. Here we derive the modified formula starting from the Gaussian lens formula. The same result was derived in [77] using a slightly different approach.

[Figure B1.1](#) shows a schematic of the entrance and exit pupils, the object and image space Principal planes, and the object and image points. In the figure,  $u$  and  $\bar{u}$  are the distances from the Principal planes to the object and image planes,  $e$  and  $\bar{e}$  are distances from the Principal planes to the entrance and exit pupils, and  $z_e$  and  $\bar{z}_e$  are the distances from the entrance and exit pupils to the object and image planes. Since the entrance and exit pupil planes are conjugates, like the object and image planes, the Gaussian lens formula holds as follows:

$$-\frac{1}{u} + \frac{1}{\bar{u}} = \frac{1}{f}, \quad (B1.1)$$

and

$$-\frac{1}{e} + \frac{1}{\acute{e}} = \frac{1}{f} . \quad (B1.2)$$



**Figure B1.1** Schematic of imaging through a lens. The figure shows the object ( $y$ ) and its image ( $\acute{y}$ ), the object space principal plane ( $H$ ) and the image side principal plane ( $\acute{H}$ ), the entrance ( $E$ ) and exit ( $\acute{E}$ ) pupils, and the associated distances along the optical axis.

The transverse magnification  $m_t$  between the object and image planes is given as:

$$m_t = \frac{\acute{u}}{u} . \quad (B1.3)$$

For images that are *real* and inverted, the transverse magnification  $m$  is numerically negative since the directed distance  $u$  is numerically negative, and  $\acute{u}$  is numerically positive.

The pupil magnification  $m_p$  is defined as the ratio of the exit pupil diameter to the entrance pupil diameter. It is also the ratio between the exit pupil and entrance pupil distances (measured from the principal planes), just like the transverse magnification between any conjugate planes:

$$m_p = \frac{\acute{e}}{e} . \quad (B1.4)$$

Equating Eqs. (B1.1) and (B1.2), we obtain

$$\frac{u - e}{eu} = \frac{\acute{u} - \acute{e}}{\acute{e}\acute{u}} .$$

Further, substituting  $z_e = u - e$  and  $\acute{z}_e = \acute{u} - \acute{e}$  in the above equation, and using Eqs. (B1.3) and (B1.4), we obtain a relationship between pupil magnification, transverse magnification, the object and image plane distances (specified with respect to the pupils) as:

$$\boxed{\frac{\acute{z}_e}{z_e} = \left(\frac{\acute{e}}{e}\right) \left(\frac{\acute{u}}{u}\right) = m_p m_t} \quad (B1.5)$$

Substituting  $u = z_e + e$  and  $\acute{u} = \acute{z}_e + \acute{e}$  in Eq. (B1.1) and equating with Eq. (B1.2) yields

$$\frac{1}{\acute{z}_e + \acute{e}} - \frac{1}{z_e + e} = \frac{1}{\acute{e}} - \frac{1}{e},$$

which after cross-multiplication and cancellations of common terms produces

$$z_e \acute{z}_e (e - \acute{e}) + e^2 \acute{z}_e - \acute{e}^2 z_e = 0.$$

Dividing throughout by  $z_e \acute{z}_e e \acute{e}$ , and substituting  $\frac{\acute{e}}{e}$  by the pupil magnification  $m_p$ , and  $\left(\frac{e - \acute{e}}{e \acute{e}}\right)$  by  $\frac{1}{f}$

we obtain:

$$\boxed{-\frac{1}{m_p z_e} + \frac{m_p}{\acute{z}_e} = \frac{1}{f}} \quad (B1.6)$$

where,

$m_p$	Pupil magnification.
$z_e$	Directed distance from the entrance pupil to the object plane.
$\acute{z}_e$	Directed distance from the exit pupil to the image plane.
$f$	Focal length.

Note that Eq. (B1.6) is valid even if the  $z_e$  and  $\acute{z}_e$  denote distances from the principal planes, provided we let  $m_p = 1$ . This outcome is indeed consistent with geometric optics theory, according to which the magnification between the principal planes is unity. In fact, Eq. (4.14) is more general than the Gaussian Lens formula in that it relates a pair of conjugate planes with any

other pair of conjugate planes for which the transverse magnification (between the planes) is known. When one of the pairs happen to be the principal planes ( $H$  and  $\tilde{H}$ ) between which the magnification is one, we obtain the Gaussian Lens formula.

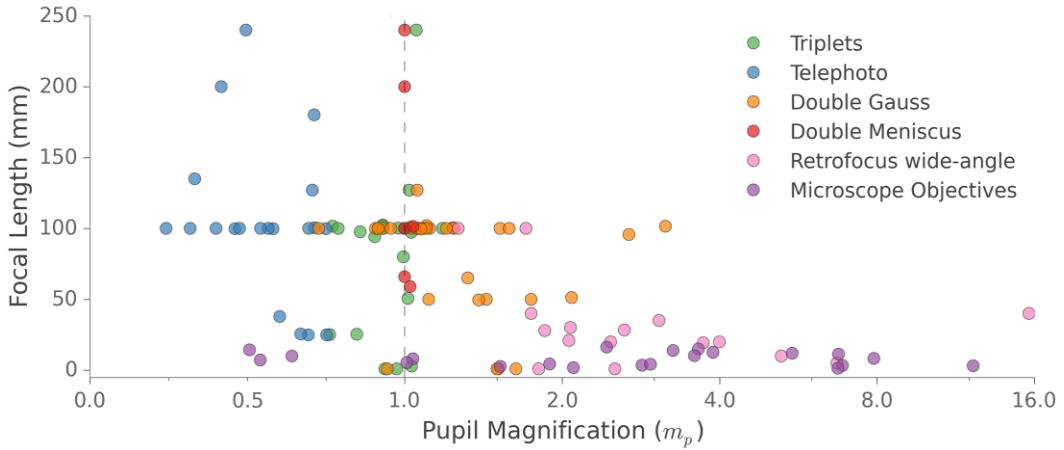
Finally, we also obtain the equations for computing the entrance and exit pupil distances from the respective principal planes by substituting Eq. (B1.4) into Eq. (B1.2) as

$$e = f \left( 1 - \frac{1}{m_p} \right), \quad \text{and} \quad (B1.7)$$

$$\epsilon = f(m_p - 1).$$

## **Appendix B.2 A brief account on the significance of pupil magnification**

Although a pupil magnification close to one is a desirable property from the point of view of distortion in the presence of orientation misalignment, it seems to be hardly a critical design choice for most practical lenses except for those used for Scheimpflug photography as evident in plot of pupil magnifications in [Figure B2.1](#). In addition, the figure shows that telephoto lenses have pupil magnification less one, and retrofocus wide-angle lenses have pupil magnification greater than one. The telephoto lenses employ a negative focal length group near the sensor plane to accommodate a long focal length lens into a compact body. Consequently, the exit pupil height (the image of the limiting aperture at the image side) is smaller compared to the entrance pupil height (the image of the limiting aperture at the object side). Therefore, in telephoto lenses the pupil magnification less than one. On the other hand, a negative focal length group is placed at the front in short focal length, retrofocus lenses to create space between the lens and sensor which results in a larger exit pupil height compared to the entrance pupil height. Thus, the pupil magnification of retrofocus lenses are greater than one.



**Figure B2.1** Pupil magnification  $m_p$  in a wide variety of lenses that form *real* images. The figure demonstrates the absence of any correlation between pupil magnification and focal length. In addition, only 20 in the sample 120 (or one in six) lenses have pupil magnification in the range  $1 \pm 0.05$ . Over 90% of all lenses have pupil magnification greater than 0.5. We obtained the samples from the Zemax Zebase library, which is a comprehensive catalogue of well-designed professional lenses.

---

The F-number (or F/#) is the ratio of the effective focal length  $f_E$  (distance between the image side Principal plane to the image plane with the object at infinity) to the paraxial entrance pupil diameter  $D_e$  [96]:

$$F/\# = \frac{f_E}{D_e} . \quad (B2.1)$$

This infinite conjugate F/# is commonly specified by lens manufacturers on lens bodies.

The pupil magnification  $m_p$  is the ratio of the exit pupil diameter  $\hat{D}_e$  to the entrance pupil diameter  $D_e$ :

$$m_p = \frac{\hat{D}_e}{D_e} . \quad (B2.2)$$

Substituting  $D_e = m_p \hat{D}_e$  from Eq. (B2.2), (B1.2) into Eq. (B2.1), we obtain

$$F/\# = \frac{m_p f_E}{\hat{D}_e} . \quad (B2.3)$$

When the object is at infinity, the distance between the image plane and the exit pupil is obtained from Eq. (B1.6) as

$$z'_e = m_p f_E \quad (\text{for } z_e = \infty). \quad (B2.4)$$

Substituting  $z'_e$  in place of  $m_p f_E$  in Eq. (B2.3) yields the alternative and equivalent definition for F-number—*as the ratio of the exit-pupil-to-image-plane distance to the exit pupil diameter*:

$$F/\# = \frac{z'_e}{D_e} \quad (\text{for } z_e = \infty), \quad (B2.5)$$

where  $z'_e$  is the distance from the exit pupil to the image plane, and  $D_e$  is the diameter of the exit pupil.

The F-number along with the wavelength  $\lambda$  determines the diffraction limited spatial resolution of optical imaging systems at the image plane as given by the equation [11,13]:

$$r = \frac{1}{\lambda F/\#} \quad \text{lp/mm}. \quad (B2.6)$$

For finite conjugate imaging, the object-plane-to-entrance-pupil distance decreases concomitant with an increase in exit-pupil-to-image-plane distance. This increase in the image plane distance effectively increases the F-number. The expression for the effective F-number in terms of the pupil magnification  $m_p$  is derived next.

Substituting  $z_e = \frac{z'_e}{m_p m_t}$  (where  $m_t$  is the transverse magnification) from Eq. (B1.5) into Eq. (B1.6), followed by simple algebraic steps yields:

$$z'_e = f(m_p - m_t). \quad (B2.7)$$

We have established that the F-number (for infinite conjugate) is the ratio of the exit-pupil-to-image-plane distance to the exit pupil diameter. To obtain the effective F-number at finite conjugates, we substitute Eq. (B2.7), the expression for the image-plane distance for finite conjugate imaging, into Eq. (B2.5):

$$F/\#_{eff} = \frac{f(m_p - m_t)}{\tilde{D}_e} . \quad (B2.8)$$

Further, substituting  $\tilde{D}_e = m_p D_e$  (Eq. (B2.2)) and replacing  $f/D_e$  with  $F/\#$  we obtain:

$$F/\#_{eff} = F/\# \left( 1 - \frac{m_t}{m_p} \right) \quad (B2.9)$$

where  $m_t$  is the transverse magnification (numerically negative for *real* images). Eq. (B2.10)(B2.9) has been derived by Mahajan (Eq. 2-80, [97]).

Now, we can obtain a more accurate equation for diffraction limited spatial resolution that is equally valid for both finite and infinite conjugate imaging by substituting Eq. (B2.9) into Eq. (B2.6)

$$r = \frac{1}{\lambda F/\# \left( 1 - \frac{m_t}{m_p} \right)} \text{ lp/mm}$$

(B2.10)

where  $\lambda$  is the wavelength,  $F/\#$  is the standard F-number defined for infinite conjugate imaging,  $m_p$  is the pupil magnification, and  $m$  is the transverse magnification (numerically negative for *real* images). When the object is at infinity,  $m_t = 0$  and Eq. (B2.10) reduces to the optical resolution expression for infinite conjugate imaging.

## APPENDIX C

### Appendix C.1 Distribution of light near focus (3D PSF) for imaging between parallel planes

The coherent PSF in image coordinates  $(u, v, z_i)$  for a point source located at  $(\xi, \eta, z_o)$  and pupil,  $P$ , at  $(x, y, 0)$  is given as [11]:

$$h(u, v, z_i; \xi, \eta, z_o) = A \iint_{-\infty}^{\infty} P(x, y) e^{-jk\frac{W_d}{2}(x^2+y^2)} e^{-j\frac{k}{z_i}[(u-M\xi)x+(v-M\eta)y]} dx dy \quad (C1.1)$$

where,  $M = -\frac{z_i}{z_o}$ ;  $k = \frac{2\pi}{\lambda}$ ;  $A = -\frac{1}{\lambda^2 z_o z_i} e^{jk z_i} e^{j\frac{k}{2z_o}(\xi^2+\eta^2)} e^{j\frac{k}{2z_o}(u^2+v^2)}$ ;  $W_d = -\left(\frac{1}{z_o} + \frac{1}{z_i} - \frac{1}{f}\right)$  is a

measure of defocus<sup>3</sup> such that  $W_d = 0$  when  $z_i = z_g$ . It can also be expressed as  $W_d = \frac{\delta z}{z_g(z_g+\delta z)}$ .

For a rotationally symmetric circular aperture of radius  $a$ , and after dropping the pure phase factors that are not important for intensity calculations, Eq. (C1.1) reduces, in polar coordinates to:

$$h(r, z_i, z_o) = \frac{2\pi a^2}{\lambda^2 z_o z_i} \int_0^1 P_r(\rho) J_0\left(\frac{k\rho a r}{z_i}\right) e^{-jk\frac{W_d}{2}\rho^2 a^2} \rho d\rho \quad (C1.2)$$

where,  $J_0$  is the Bessel function of first kind, order zero, and

$$x = \rho a \cos \theta \quad u = r \cos \phi$$

$$y = \rho a \sin \theta \quad v = r \sin \phi$$

$$0 \leq \rho \leq 1$$

Eq. (C1.2) is a function of the radial and defocus parameters  $r$  and  $\delta z$  respectively, which for a uniform aperture-transmittance function reduces to:

---

<sup>3</sup> The relation between  $W_d$  and the Seidel aberration coefficient for defocus,  $W_{020}$ , can be expressed as  $W_{020} = a^2 W_d / 2$ .

$$h(r, \delta z) = \frac{2\pi a^2}{\lambda^2 z_o (z_g + \delta z)} \int_0^1 J_o \left( \frac{k\rho a r}{z_g + \delta z} \right) e^{-jk \frac{W_d}{2} \rho^2 a^2} \rho d\rho \quad (C1.3)$$

Further, the normalized intensity distribution plotted in [Figure 1.2 \(b\)](#), which can be solved in terms of Lommel functions following the steps shown by Born et al. [13], is obtained as:

$$I_n(r, \delta z) = \frac{|h(r, \delta z)|^2}{I_o}, \text{ where } I_o = \left[ \frac{2\pi a^2}{\lambda^2 z_o z_g} \right]^2 \quad (C1.4)$$

## REFERENCES

