Depth from Refraction Using a Transparent Medium with Unknown Pose and Refractive Index

Zhihu Chen · Kwan-Yee K. Wong · Yasuyuki Matsushita · Xiaolong Zhu

Abstract In this paper, we introduce a novel method for depth acquisition based on refraction of light. A scene is captured directly by a camera and by placing a transparent medium between the scene and the camera. A depth map of the scene is then recovered from the displacements of scene points in the images. Unlike other existing depth from refraction methods, our method does not require prior knowledge of the pose and refractive index of the transparent medium, but instead can recover them directly from the input images. By analyzing the displacements of corresponding scene points in the images, we derive closed form solutions for recovering the pose of the transparent medium and develop an iterative method for estimating the refractive index of the medium. Experimental results on both synthetic and real-world data are presented, which demonstrate the effectiveness of the proposed method.

Keywords Refraction · Pose estimation · Refractive index · Depth · Reconstruction

1 Introduction

Depth from refraction is a depth acquisition method based on refraction of light. A scene is captured several times by a fixed perspective camera, with the first image captured directly by the camera and the others by placing a transparent medium between the scene and the camera. The depths of the scene points are then recovered from their displacements in the images.

Depth from refraction approach has various advantages over other existing 3D reconstruction approaches. First, unlike multiple-view stereo methods, it does not require calibrating the relative rotations and translations of the camera as the viewpoint is fixed. Besides, a fixed viewpoint also makes the correspondence-problem much easier as the projections of a 3D point remain similar across images. Second, unlike depth from defocus methods which require expensive lenses with large apertures to improve depth sensitivity, the accuracy of depth from refraction can be improved by increasing either (1) the angle between the viewing direction of a 3D point and the surface normal of the medium; (2) the refractive index of the medium; or (3) the thickness of the refractive medium. Third, unlike depth from diffusion methods, which require placing a diffuser with a known orientation close to the object being measured, depth from refraction allows the transparent medium being placed flexibly between the scene and the camera.

Existing depth from refraction methods often require elaborate setup and tedious offline calibration for accurately knowing the pose and refractive index of the transparent medium. These greatly prohibit the applicability of the approach. In this paper, we introduce a novel method for depth from refraction which is more usable in various scenarios. Our method requires neither a careful hardware setup nor any offline calibration[1]. By simply putting a transparent medium between a camera and the scene, our method automatically estimates the pose and refractive index of the transparent medium as well as a depth map of the scene.

In our method, a scene is captured twice by a fixed perspective camera, with the first image (referred to as the

Z. Chen (✉) · K.-Y. K. Wong · X. Zhu
Department of Computer Science, The University of Hong Kong, Hong Kong, Hong Kong
e-mail: zhchen04@gmail.com

Y. Matsushita
Microsoft Research Asia, Beijing , China

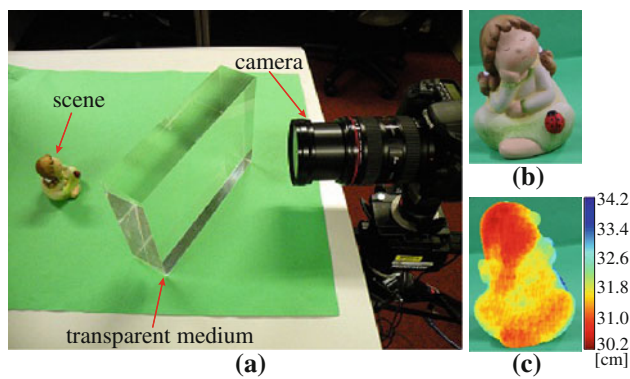[1] Save for the intrinsic parameters of the camera

Fig. 1 (**a**) Experimental setup. (**b**) A direct image of the scene. (**c**) A reconstructed depth map of the scene

direct image) captured directly by the camera and the second (referred to as the refracted image) by placing a transparent medium with two parallel planar faces between the scene and the camera (see Fig. 1). By analyzing the displacements of the scene points in the images, we derive closed form solutions for recovering both the pose of the transparent medium and the depths of the scene points. Given a third image captured with the transparent medium placed in a different pose, we further develop an iterative method for recovering also the refractive index of the medium.

2 Related Work

Depth acquisition has a long history in computer vision. Based on the number of viewpoints required, existing methods can be broadly classified into multiple-view and multiple-exposure approaches. Multiple-view methods exploit stereo information to recover the depth of a scene (Scharstein and Szeliski 2002). The location of a 3D point can be estimated by finding and triangulating correspondences across images.

Instead of moving the camera to change the viewpoints, multiple-exposure methods record the scene by changing the imaging process. Depth from defocus methods obtain depth by exploiting the fact that depth information is contained in an image taken with a limited field of depth: objects at a particular distance are focused in the image, while objects at other distances are blurred by different degrees depending on their distances. Pentland(1987) estimated a depth map of a scene by measuring the degree of defocus using one or two images. In 1988, Subbarao and Gurumoorthy proposed a simpler and more general method to recover depth by measuring the degree of blur of an edge. Surya and Subbarao (1993) used simple local operations on two images taken by cameras with different aperture diameters for determining depth. Zhou et al. (2009) pointed out that the accuracy of depth is restricted by the use of a circular aperture. They proposed a comprehensive framework to obtain an optimized

pair of apertures. Note that all of the aforementioned methods require large apertures to improve depth sensitivity. Recently, Zhou et al. (2010) proposed a depth from diffusion method. Their method requires placing a diffuser with known orientation near the scene. They showed that while depth from diffusion is similar in principle to depth from defocus, it can improve the accuracy of depth obtained with a small lens by increasing the diffusion angle of a diffuser.

Our work is more closely related to Lee and Kweon (2000), Maas (1995), Nishimoto and Shirai (1987), Gao and Ahuja (2004), (2006), Shimizu and Okutomi (2006), (2007). In Lee and Kweon (2000), Lee and Kweon obtained the geometry of an object using a transparent biprism. In Maas (1995), Maas estimated the depth of a scene in the field of multimedia photogrammetry using a transparent planar plate under the constraint that the faces of the planar plate were parallel to the image plane. In Nishimoto and Shirai (1987), Nishimoto and Shirai removed the constraint and proposed a system using a tilted parallel planar plate to achieve single camera stereo. Their system is based on the assumption that positioning a tilted parallel planar plate between the camera and the scene results in a new pinhole camera, which, however, is invalid in general. In Gao and Ahuja (2004), Gao and Ahuja proposed a more accurate single camera stereo system with a parallel planar plate in an arbitrary orientation. They estimated the pose and refractive index of the medium in an extra step using a calibration pattern. In Gao and Ahuja (2006), Gao and Ahuja rotated a parallel planar plate around the principal axis of a camera while keeping its tilt angle fixed. This method avoids repeated calibration of the medium pose. Nevertheless, it requires a complex device to guarantee the new dimension of control, which is hard to achieve in practice. In Shimizu and Okutomi (2006) and Shimizu and Okutomi (2007), Shimizu and Okutomi proposed *reflection stereo* which records the scene from a fixed viewpoint with and without the reflection light paths. The two light paths create individual images, and from these images, their method estimates depth by triangulation. Their setup uses either reflective or refractive medium for the implementation. However, their method requires a complex calibration setup as described in Shimizu and Okutomi (2008).

Similar to many of the above work, our method also uses a transparent parallel planar medium. However, unlike the aforementioned methods, our method requires neither the medium plane being parallel to the image plane, nor a careful calibration of the pose and refractive index of the medium using a calibration pattern. In this paper, we describe several extensions to our previous work on depth from refraction (Chen et al. 2011). Firstly, we present a better derivation for the closed form solution to depth recovery of a 3D point. Secondly, we discuss in detail why total internal reflection will never occur in our setup. Thirdly, we make an analogy between the configuration of our method and epipolar

geometry. Fourthly, we discuss the wavelength dependency of the refractive index and how it affects the accuracy in depth estimation. Finally, we present an experiment to compare our method against a standard stereo method.

Shape recovery of transparent objects (also referred to as refractive objects) has also attracted the attentions of many researchers. In Murase (1992), Murase used optical flow to recover the shape of a nonrigid transparent object from the apparent motion of an observed pattern under a number of simplifying assumptions. In Zhang and Cox (1994), Zhang and Cox developed a method to measure the slopes of a water surface covered with short waves. However, their method relies on special optics and very complicated experimental settings, making it not very practical. In Agarwal et al. (2004), Agarwal et al. presented a method for recovering the shape of a rigid and stationary refractive object from a video sequence of the apparent motion of the background behind the object. In Morris and Kutulakos (2011), Morris and Kutulakos pointed out that two viewpoints are sufficient to obtain the 3D positions and normals of points on an unknown refractive surface even if the refractive index is unknown. All of the aforementioned methods target at reconstructing refractive surfaces. This paper, on the other hand, exploits a refractive object to obtain the depth of a scene.

The rest of the paper is organized as follows. Section 3 briefly reviews the theory on refraction of light. Section 4 describes our proposed method in detail. We first derive a closed form solution for recovering the depth of a scene in the case where the parallel planar faces of the transparent medium are parallel to the image plane. We further show that this closed form solution can be extended to the case where the parallel planar faces of the medium are not parallel to the image plane but have a known orientation. Moreover, we develop a novel method for recovering the orientation of the parallel planar faces of the medium from one direct image and one refracted image. Finally, by capturing a second refracted image with the transparent medium under a different pose, we develop an iterative method for recovering the refractive index of the medium. Section 5 first explains why total internal reflection will never occur in the proposed setup. It then presents an analogy between the proposed setup and stereo vision, and introduces a geometric constraint which helps the finding of correspondences across images. Finally, it discusses the wavelength dependency of the refractive index and its effect on depth estimation. Experimental results on both synthetic and real data are presented in Sect. 6, followed by conclusions in Sect. 7.

3 Refraction of Light

Refraction of light refers to the change in the direction of a light ray due to a change in its speed. This is commonly
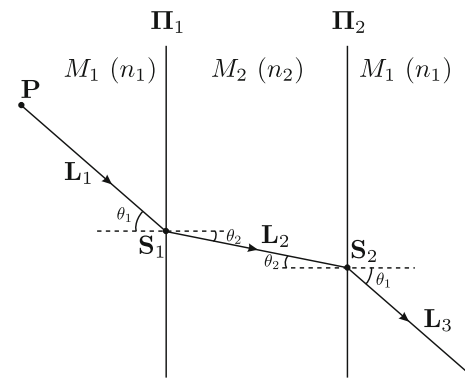

Fig. 2 When a light ray passes from one medium to another with different speeds, it bends according to Snell's law. This phenomenon is known as refraction of light

observed when a light ray passes from one medium to another (e.g., from air to water). The *refractive index* of a medium is defined as the ratio of the velocity of light in vacuum to the velocity of light in the said medium. Consider a light ray L_1 originated from a point P passing from a medium M_1 to another medium M_2 (see Fig. 2). Let the refractive indices of M_1 and M_2 be n_1 and n_2 respectively, and the interface between the two media be a plane denoted by Π_1. Suppose L_1 intersects Π_1 at S_1 with an angle of incidence θ_1. After entering M_2, L_1 changes its direction and results in a refracted ray L_2 with an angle of refraction θ_2. By Snell's law, the incident ray L_1, the surface normal at S_1 and the refracted ray L_2 are coplanar, and the angle of incidence θ_1 and the angle of refraction θ_2 are related by

$$n_1 \sin \theta_1 = n_2 \sin \theta_2. \tag{1}$$

After traveling for some distance in M_2, L_2 leaves M_2 and enters M_1 again. Let the interface between M_2 and M_1 be a plane denoted by Π_2 which is parallel to Π_1. Suppose L_2 intersects Π_2 at S_2 and after reentering M_1, it changes its direction and results in a refracted ray L_3. Since Π_1 and Π_2 are parallel, it is easy to see that the angle of incidence for L_2 is θ_2. It follows from Snell's law that the angle of refraction for L_3 is θ_1, and L_1, L_2, L_3 and the surface normals of Π_1 and Π_2 are coplanar, with L_1 being parallel to L_3. Hence, the refraction plane of P (formed by L_1, L_2 and L_3) is perpendicular to both Π_1 and Π_2.

In practice, when a light ray passes from one medium to another, it will not only be transmitted (with refraction) through the second medium, but will also be partially reflected back to the original medium (see Fig. 3). When the refractive index of the first medium is higher than that of the second (e.g., from water to air), a phenomenon known as *total internal reflection* might occur in which the light ray will be totally reflected back to the first medium without transmitting through the second medium. Consider again a light ray passing from M_1 to M_2, and let $n_1 > n_2$ (see Fig. 3). Since
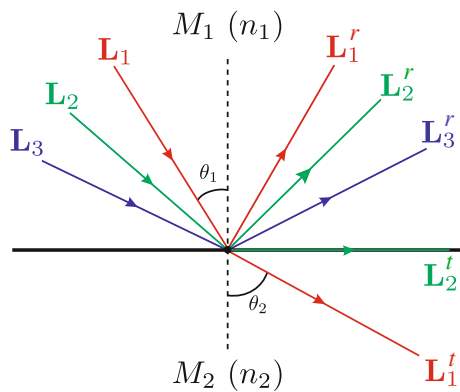
Fig. 3 Total internal reflection. $\mathbf{L}_i$, $\mathbf{L}_i^r$ and $\mathbf{L}_i^t$, $i \in \{1, 2, 3\}$, denote the incident, reflected and refracted rays respectively

the sine function is a strictly increasing function for angle between 0° and 90°, it follows from Eq. (1) that $\theta_2 > \theta_1$. As θ_1 increases, θ_2 increases. When θ_1 reaches the *critical angle* $\theta_c = \arcsin \frac{n_2}{n_1}$, θ_2 becomes 90° and the refracted ray travels along the interface between M_1 and M_2. When the angle of incidence θ_1 is greater than the critical angle θ_c, total internal reflection occurs and the light ray will be completely reflected back to M_1 without transmitting through M_2.

4 Depth from Refraction

In this section, we will derive a closed form solution for recovering the depth of a scene from the displacements of scene points observed in two images due to refraction of light. As mentioned in Sect. 1, a scene will be captured twice by a fixed perspective camera, with the first image (referred to as the direct image) captured directly by the camera and the second (referred to as the refracted image) by placing a transparent medium between the scene and the camera. We assume the intrinsic parameters of the camera are known, and the transparent medium consists of two parallel planar faces through which light rays originate from scene points enter and leave the medium before reaching the camera.

4.1 Medium Surface ‖ Image Plane

Consider a 3D point $\mathbf{P}$ being observed by a camera centered at $\mathbf{O}$ (see Fig. 4). Let the direct projection of $\mathbf{P}$ on the image plane be a point $\mathbf{I}$. Suppose now a transparent medium M with two parallel planar faces is placed between $\mathbf{P}$ and the camera in such a way that the two parallel planar faces are parallel to the image plane. Due to refraction of light, $\mathbf{P}$ will no longer project to $\mathbf{I}$. Let $\mathbf{I}'$ be the new image position for the projection of $\mathbf{P}$. By considering the orthographic projection of the line $\mathbf{PO}$ on the image plane, and relating it to the orthographic projections of $\mathbf{PS}_1$, $\mathbf{S}_1\mathbf{S}_2$ and $\mathbf{S}_2\mathbf{O}$ on the image plane, we have
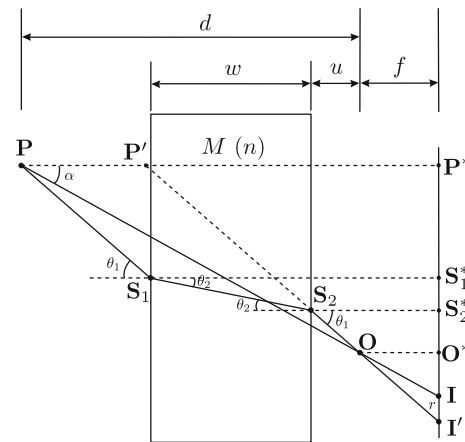

Fig. 4 Special case of shape from refraction in which the parallel planar faces of the medium are parallel to the image plane. The depth of a 3D point can be recovered in closed form

$$|\mathbf{P}^*\mathbf{O}^*| = |\mathbf{P}^*\mathbf{S}_1^*| + |\mathbf{S}_1^*\mathbf{S}_2^*| + |\mathbf{S}_2^*\mathbf{O}^*|. \tag{2}$$

Furthermore, from the geometry shown in Fig. 4, we have

$$\begin{aligned}
|\mathbf{P}^*\mathbf{O}^*| &= d \tan \alpha, \\
|\mathbf{P}^*\mathbf{S}_1^*| &= (d - w - u) \tan \theta_1, \\
|\mathbf{S}_1^*\mathbf{S}_2^*| &= w \tan \theta_2, \\
|\mathbf{S}_2^*\mathbf{O}^*| &= u \tan \theta_1.
\end{aligned} \tag{3}$$

where d is the depth of $\mathbf{P}$, α is the angle between the visual ray of $\mathbf{P}$ and the principal axis of the camera (also referred to as the *viewing angle* of $\mathbf{P}$), w is the thickness of the medium M, u is the shortest distance between $\mathbf{O}$ and M, and θ_1 and θ_2 are the angle of incidence/refraction and angle of refraction/incidence as the ray originated from $\mathbf{P}$ enters/leaves M. Substituting Eq. (3) into Eq. (2) gives

$$d \tan \alpha = (d - w - u) \tan \theta_1 + w \tan \theta_2 + u \tan \theta_1, \tag{4}$$

Rearranging Eq. (4) gives

$$d = w \frac{\tan \theta_1 - \tan \theta_2}{\tan \theta_1 - \tan \alpha}. \tag{5}$$

By Snell's law, we have

$$\sin \theta_1 = n \sin \theta_2, \tag{6}$$

where n is the refractive index of M and the refractive index of air can be approximated to one. Let r be the distance between $\mathbf{I}$ and $\mathbf{I}'$. It can be expressed in terms of the focal length f of the camera and the angles θ_1 and α as

$$r = f(\tan \theta_1 - \tan \alpha). \tag{7}$$

From Eqs. (6) and (7), we can obtain the following expressions for $\tan \theta_1$ and $\tan \theta_2$:

$$\tan \theta_1 = \frac{r}{f} + \tan \alpha, \tag{8}$$

$$\tan \theta_2 = \sqrt{\frac{\tan^2 \theta_1}{n^2 + (n^2 - 1)\tan^2 \theta_1}}. \tag{9}$$

Finally, substituting Eqs. (8) and (9) into Eq. (5) gives

$$d = w \left(1 + \frac{f}{r} \tan \alpha \right)$$
$$\times \left(1 - \sqrt{\frac{1}{n^2 + (n^2 - 1)(\frac{r}{f} + \tan \alpha)^2}} \right). \tag{10}$$

From Eq. (10), we have the following three observations:

1. d does not depend on u. This is very important in practice as it implies that the depth of a scene point can be recovered without knowing the distance between the medium M and the camera;
2. d is scaled by w. It implies that without knowing the thickness w of the medium, the depth of the scene can still be reconstructed up to an unknown scale.
3. d increases monotonically with n ($n > 1$) when all other parameters are fixed. It implies that if the refractive index used is larger than the ground truth, the recovered point will be farther away form the camera, and visa versa.

4.2 Medium Surface ∦ Image Plane

It has been shown in the previous subsection that depth can be recovered using Eq. (10) when the parallel planar faces of the transparent medium are parallel to the image plane. However, this proposed setup has two major limitations. First, it is difficult to ensure that the parallel planar faces of the medium are parallel to the image plane. Second, the depths of those 3D points whose projections are near the image center (*i.e.*, with small viewing angles) will be very sensitive to noises (see Fig. 9a). As a result, only points with large viewing angles (*i.e.*, those projected near the border of the image) can be accurately recovered. In this subsection, we will show how the closed form solution derived under the special case can be applied to recover depth in the general case where the parallel planar faces of the medium are not parallel to the image plane but have a known orientation.

Suppose the surface normal of the parallel planar faces of the medium is given by the unit vector **N**, and the viewing direction of the camera is given by the unit vector **V** (see Fig. 5). A rotation (represented by a rotation matrix **R**) about the optical center of the camera, with a rotation axis given by the cross product of **V** and **N** and a rotation angle given by the angle between **V** and **N**, will bring the image plane parallel to the parallel planar faces of the medium. Such a rotation will induce a planar homography $\mathbf{H} = \mathbf{KRK}^{-1}$, where **K** is the camera calibration matrix, that can transform the image of
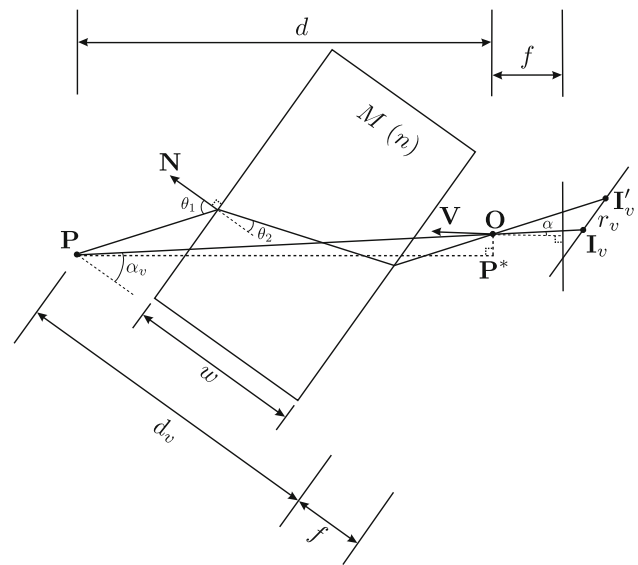

Fig. 5 When the parallel planar faces of the medium are not parallel to the image plane, the depth of a 3D point can be recovered in closed form if the orientation of the medium is known

the original camera to an image observed by the camera after rotation. The closed form solution derived in the previous subsection can then be applied to the transformed image to recover the depth d_v of a point **P** with respect to the rotated camera using the viewing angle α_v of **P** and the displacement r_v of the projections of **P** in the rotated camera. Referring to Fig. 5, we have

$$\frac{d_v}{|\mathbf{PO}|} = \cos \alpha_v, \tag{11}$$

$$\frac{d}{|\mathbf{PO}|} = \cos \alpha. \tag{12}$$

Combining Eqs. (11) and (12) gives

$$d = d_v \frac{\cos \alpha}{\cos \alpha_v}. \tag{13}$$

4.3 Recovering Pose of the Medium

It has been shown in the previous subsection that scene depth can be recovered given the orientation of the parallel planar faces of the transparent medium. In this subsection, we will show how the orientation of the parallel planar faces of the medium can be recovered directly from the displacements of the scene points in the images.

Consider a 3D point **Q** whose visual ray is perpendicular to the parallel planar faces of the medium (see Fig. 6). By construction, the visual ray of **Q** will simply pass straight through the medium without any change of direction. Hence, the projections of **Q** will be identical in both the direct image and refracted image. Let us denote this point by **J**. Without loss of generality, consider another 3D point **P**, and let
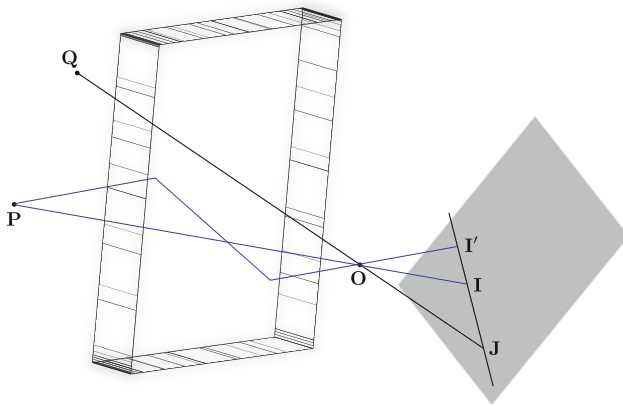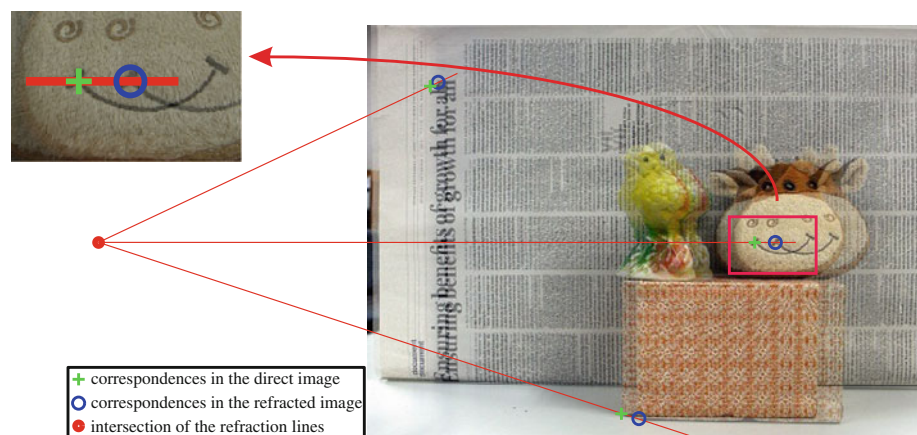
Fig. 6 Recovering the orientation of the parallel planar faces of the transparent medium. A refraction line defined by corresponding points in the direct and refracted images contains the vanishing point for the normal direction of the parallel planar faces of the medium

I and **I′** be the projections of **P** in the direct image and the refracted image respectively. Referring to Sect. 3, the refraction plane of **P** is perpendicular to the parallel planar faces of the medium. Since this plane contains both **P** and **O**, and that the ray **QO** is, by construction, also perpendicular to the parallel planar faces of the medium, it follows that **Q** must also lie on this plane. This plane intersects the image plane along a line which we refer to as a *refraction line*. It is obvious that both **J**, **I** and **I′** must lie on this line. Now consider two or more refraction lines defined by correspondences in the direct and refracted images (see Fig. 7). As all such refraction lines must contain **J**, they must intersect at **J** which in fact is the vanishing point for the normal direction of the parallel planar faces of the medium. Based on this observation, we have the following proposition:

Proposition 1 *Given the correspondences of two or more scene points in a direct image and a refracted image, the refraction lines defined by the correspondences will intersect at a single point which corresponds to the vanishing point*

for the normal direction of the parallel planar faces of the medium.

The corollary below then follows directly from Proposition 1:

Corollary 1 *The orientation of the parallel planar faces of the transparent medium can be recovered as the visual ray for the point of intersection between two or more refraction lines defined by correspondences in the direct image and the refracted image.*

4.4 Estimation of the Refractive Index

In the previous discussions, we have assumed that the refractive index of the transparent medium is known a priori. In this subsection, we will show the refractive index of the medium can be recovered from the displacements of scene points in three images.

Consider a 3D point **P**. Let **I** be its projection in a direct image, and **I′** and **I″** be its projections in two refracted images captured with the transparent medium positioned in two different poses respectively. Let d be the depth of **P** estimated from **I** and **I′** using Eqs. (10) and (13), and $d′$ be the depth of **P** estimated from **I** and **I″** using Eqs. (10) and (13). Now by equating d with $d′$, the refractive index n of the medium can be recovered. In practice, given m pairs of correspondences in three images, the refractive index of the medium can be estimated by

$$n = \arg\min_n \sum_{i=1}^{m} (d_i(n) - d_i'(n))^2. \tag{14}$$

Note that a similar minimization can also be used to estimate the refractive index when there are more than two refracted images captured with the transparent medium positioned in different poses.

Fig. 7 Recovering the orientation of the parallel planar faces of the transparent medium. Refraction lines defined by corresponding points in the direct and refracted images must intersect at the vanishing point for the normal direction of the parallel planar faces of the medium

+ correspondences in the direct image
○ correspondences in the refracted image
● intersection of the refraction lines

5 Discussions

In this section, we first show that total internal reflection will never occur in our proposed setup. We then present an analogy between our proposed setup and stereo vision, and introduce a geometric constraint which helps finding of correspondences across images. Finally, we discuss the wavelength dependency of the refractive index and how it affects the accuracy of depth estimation.

5.1 Total Internal Reflection

It has been pointed out in Sect. 3 that total internal reflection can only occur when (1) a light ray passes from one medium with a higher refractive index to one with a lower refractive index, and (2) the angle of incidence is greater than the critical angle. In our setup, light rays travel through a composite air-glass-air medium (see Fig. 4). Obviously, total internal reflection will not occur when a light ray passes from air to glass as the refractive index of air is lower than that of glass. From Eq. (6), the angle of refraction θ_2 is given by $\arcsin(\sin\theta_1/n)$. Since $0° \leq \theta_1 < 90°$ and the sine function is a strictly increasing function for the range of θ_1, $\theta_2 < \arcsin(1/n)$. Note that $\arcsin(1/n)$ is in fact the critical angle for a light ray passing from glass to air. It follows that when the light ray travels through the glass medium and re-enters air, the angle of incidence θ_2 will always be less than the critical angle, and hence total internal reflection will never occur.

5.2 Depth from Refraction Versus Stereo Vision

Although the image pairs used in depth from refraction look very similar to those used in stereo vision, there exists a major difference in the image formation process: the refracted image cannot be described by a simple pinhole camera model. Consider a light ray $\mathbf{L}$ originated from a point $\mathbf{P}$ which is perpendicular to the medium surface, and let $\mathbf{L}_1$ be a light ray originated from a point $\mathbf{P}_1$ which intersects $\mathbf{L}$ at a point $\mathbf{O}_1$, and $\mathbf{L}_2$ be a light ray originated from a point $\mathbf{P}_2$ which intersects $\mathbf{L}$ at a point $\mathbf{O}_2$ (see Fig. 8). Suppose after traveling through the composite air-glass-air medium, the light paths of these three light rays intersect at the camera center $\mathbf{O}$. The distance between $\mathbf{O}$ and $\mathbf{O}_1$ is given by

$$ds_1 = w - w\frac{\tan\theta_2}{\tan\theta_1} = w\left(1 - \frac{\sqrt{1 - \sin^2\theta_1}}{\sqrt{n^2 - \sin^2\theta_1}}\right). \qquad (15)$$

It can be seen that ds_1 depends on the incident angle θ_1 of $\mathbf{L}_1$. A similar expression can be derived for the distance ds_2 between $\mathbf{O}$ and $\mathbf{O}_2$, which depends on the incident angle of $\mathbf{L}_2$. In general, $\mathbf{L}_1$ and $\mathbf{L}_2$ have different incident angles and therefore $ds_1 \neq ds_2$ (i.e., $\mathbf{O}_1$ and $\mathbf{O}_2$ do not coincide). It
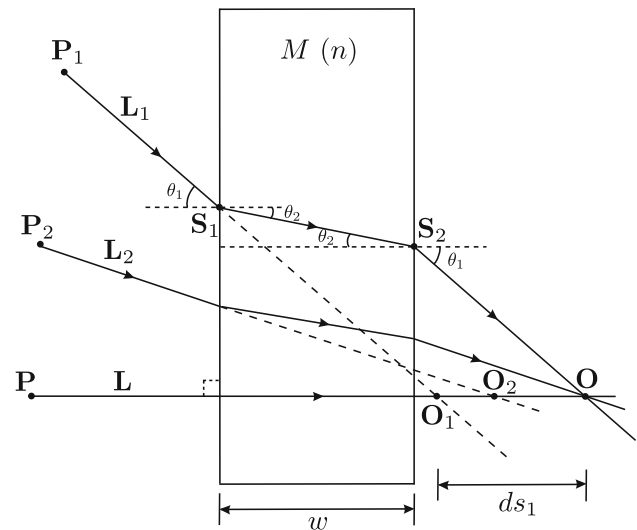

Fig. 8 The formation process of the refracted image cannot be described by a simple pinhole camera model

follows that the light rays do not intersect at a single point and therefore the image formation process of the refracted image cannot be described by a simple pinhole camera model.

Despite the differences in geometry between depth from refraction and stereo vision, there exists some very similar properties between these two approaches. First of all, the line passing through the camera center $\mathbf{O}$ and parallel to the surface normal $\mathbf{N}$ of the parallel planar faces of the transparent medium is analogous to the baseline in stereo vision. Here we call this line the *pose line* as it depends on the pose of the medium. The vanishing point of the pose line is the *focus of refraction* (previously denoted as $\mathbf{J}$ in Sect. 4.3), which is analogous to the epipole. Similar to an epipolar plane which is defined by the baseline and an arbitrary point in 3D space, a *refraction plane* is defined by the pose line and an arbitrary point in 3D space, and both the direct and refracted light paths of the point must lie on this plane. A refraction plane intersects with an image plane along a *refraction line*, just like an epipolar plane intersects with an image plane along an epipolar line.

Similar to stereo vision, depth from refraction relies on successfully establishing correspondences across images. Just in the same manner that epipolar lines can constrain the search of correspondences, refraction lines can be exploited to constrain the search of correspondences between the direct image and refracted image. As shown in Fig. 6, given the vanishing point $\mathbf{J}$ for the normal direction of the parallel planar faces of the medium and the projection $\mathbf{I}$ of a scene point in the direct image, its correspondence $\mathbf{I}'$ in the refracted image must lie on the refraction line defined by $\mathbf{J}$ and $\mathbf{I}$. Furthermore, it is obvious that $|\mathbf{JI}'| > |\mathbf{JI}|$, and $\mathbf{JI}'$ has the same direction as $\mathbf{JI}$. Based on this observation, we have the following corollary:
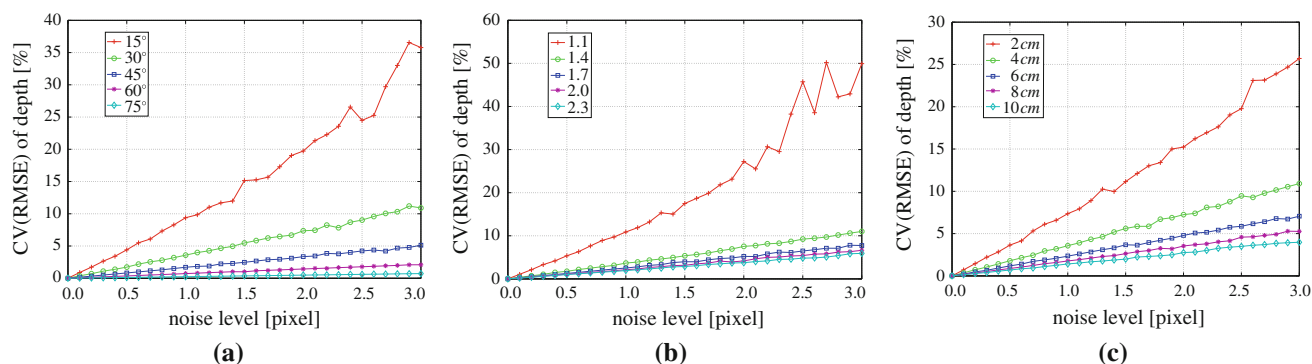
Fig. 9 Depth accuracy can be improved in three ways: (**a**) increasing the angle between the visual ray and the surface normal of the medium; (**b**) increasing the refractive index; or (**c**) increasing the thickness of the medium

Corollary 2 *Given the projection* **I** *of a scene point* **P** *in the direct image, its correspondence* **I**′ *in the refracted image must lie on the half-infinite line* **I**′$(t) = $ **I** $+ t\,($**I** $-$ **J**$)$ *where* $t \geq 0$ *and* **J** *is the the vanishing point for the normal direction of the parallel planar faces of the transparent medium.*

Having recovered the vanishing point **J** for the normal direction of the parallel planar faces of the medium from some seed correspondences, more correspondences can then be established with ease using the refraction line constraint derived from **J**.

5.3 Wavelength Dependency of the Refractive Index

An empirical relationship between the refractive index and wavelength for a particular transparent medium was proposed in the field of Physics (Sellmeier 1871), which implies that the refractive index of a medium varies with wavelength. Nevertheless, the refractive indexes of many optical materials only change slightly (typically with a maximum difference of 0.02) within the visible spectrum (i.e., with wavelength ranging from 380 to 750 nm) (Ishigure et al. 1996). Take the glass used in our real experiment as an example. The glass is made up of poly(methyl methacrylate), also known as PMMA, and has a refractive index range of [1.485, 1.503] within the visible spectrum. In Sect. 6.1, we present an experiment showing how the error of the estimated depth varies with the refractive index. Note that the maximum difference of refractive indices within the visible spectrum for the glass used in the real experiments is only 0.018. If we consider the CV(RMSE) for a refractive index error of 0.018, the error is quite small. It is also shown that although the absolute error of the estimated depth is relatively large when using a highly inaccurate refractive index (with an error of 0.15), the error becomes very small after the reconstructed model has been aligned with the ground truth by a similarity transformation (see Fig. 12). It implies that the shape of the reconstruction result has little distortion even in the case of using an

inaccurate refractive index. The shape distortion caused by different refractive indices within the visible spectrum, therefore, can be ignored.

6 Experiments

The methods described in Sect. 4 for recovering the pose and refractive index of the medium, and the depth of the scene have been implemented. Experiments on both synthetic and real data were carried out and the results are presented in the following subsections.

6.1 Synthetic Experiments

The first set of synthetic experiments was designed to evaluate the performance of depth estimation in the presence of noise. The experimental setup consisted of a 3D point being viewed by a synthetic camera with a focal length of 24 mm, and the parallel planar faces of the transparent medium were parallel to the image plane. In our synthetic experiment, the refracted and direct images of the point are computed by forward projecting the point to the image plane with and without the transparent medium, respectively (see more details in Appendix). To simulate the effect of noise, normally distributed random noise was added independently to the pixel coordinates of the projections of the 3D point with a standard deviation $\sigma \in [0.0\ 3.0]$ (in pixels). For each noise level, 1000 independent trials were carried out. As an error measurement, the coefficient of variation of root mean square error, referred to as CV(RMSE), was computed by normalizing the root mean square error of the depth estimates by the mean depth. In the first experiment, the thickness of the medium was 4 cm and its refractive index was 1.4. Experiments were carried out for distinct viewing angles. Figure 9a shows a plot of CV(RMSE) of depth against different noise levels for different viewing angles. It can be seen that for a particular noise level, depth accuracy can be
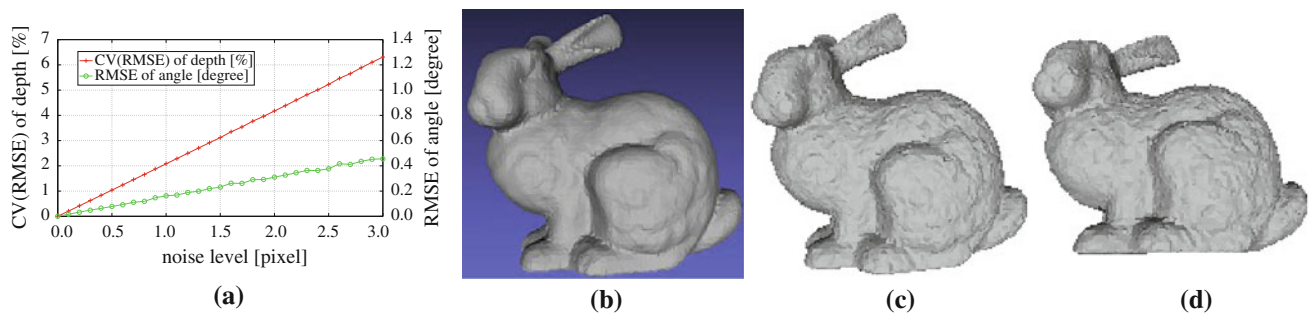
Fig. 10 Reconstruction of the bunny model using a direct image and a refracted image. (**a**) Estimation errors for the orientation of the parallel planar faces of the medium and for the depths of points on the bunny model. (**b**) Original bunny model. (**c**), (**d**) Two views of the reconstructed bunny model
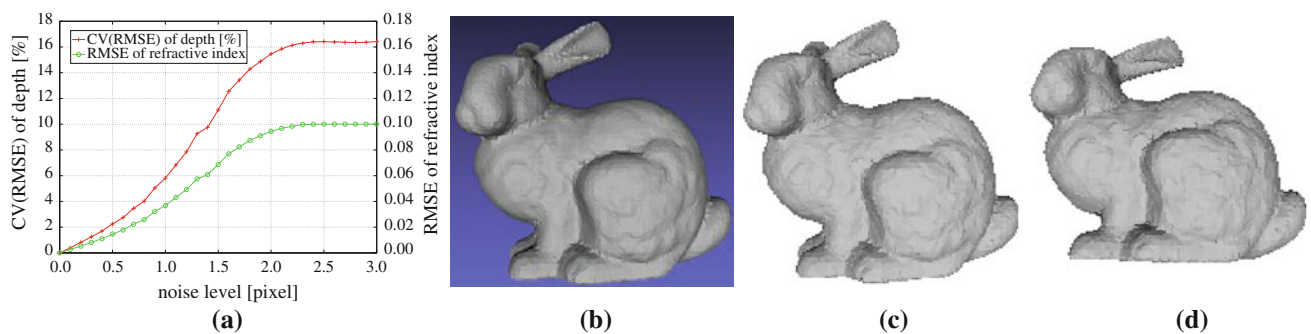

Fig. 11 Reconstruction of the bunny model using a direct image and two refracted images captured with the transparent medium in two different poses. (**a**) Estimation errors for the refractive index of the medium and for the depths of points on the bunny model. (**b**) Original bunny model. (**c**), (**d**) Two views of the reconstructed bunny model

improved by increasing the viewing angle in the case when the medium surfaces are parallel to the image plane. In the general case, it implies that depth accuracy can be improved by increasing the angle between the visual ray and the surface normal of the medium. In the second experiment, the viewing angle was 30° and thickness of the medium was 4 cm. Experiments were carried out for different refractive indices. Figure 9b shows a plot of CV(RMSE) of depth against different noise levels for different refractive indices. It can be seen that for a particular noise level, depth accuracy can be improved by increasing the refractive index of the medium. In the third experiment, the viewing angle was 30° and refractive index of the medium was 1.4. Experiments were carried out for different thicknesses of the medium. It is shown in Fig. 9c that for a particular noise level, depth accuracy can be improved by increasing the thickness of the medium.

In the synthetic experiment of estimating the orientation of the parallel planar faces of the medium, a bunny model was captured using a synthetic camera and a 4 cm thick transparent medium with a refractive index of 1.4. To simulate the effect of noise, normally distributed random noise with $\sigma \in [0.0\ 3.0]$ (in pixels) was added independently to the pixel coordinates of 8171 points on the bunny model. The image resolution of the bunny model was 1500×1500. For

each noise level, 1000 independent trials were carried out. The orientation of the parallel planar faces of the medium and the 3D coordinates of all the points were obtained using correspondences in the direct image and the refracted image. Figure 10a shows the RMSE (in terms of angle) of the estimates for the surface normal of the parallel planar faces, as well as CV(RMSE) of depth against different noise levels. Under a noise level of 1.0 pixel, the RMSE (in terms of pixel) of distance between the estimated vanishing point and the ground truth is 16.93 pixels. Compared to the distance of 4352.7 pixels between the vanishing point and the image center, the error is small and the refraction line constraint still holds, which is proved by the fact that the RMSE of distance between correspondences in the refracted image and the refraction lines is 1.26 pixels. The reconstruction result is shown in Fig. 10c, d.

In the synthetic experiment of estimating the refractive index of the medium, a third image of the bunny model was captured with the transparent medium positioned in a different pose. The orientations of the parallel planar faces of the medium were estimated using correspondences between the direct image and each of the two refracted images. The refractive index of the medium and the 3D coordinates of all the points were then obtained using correspondences in all
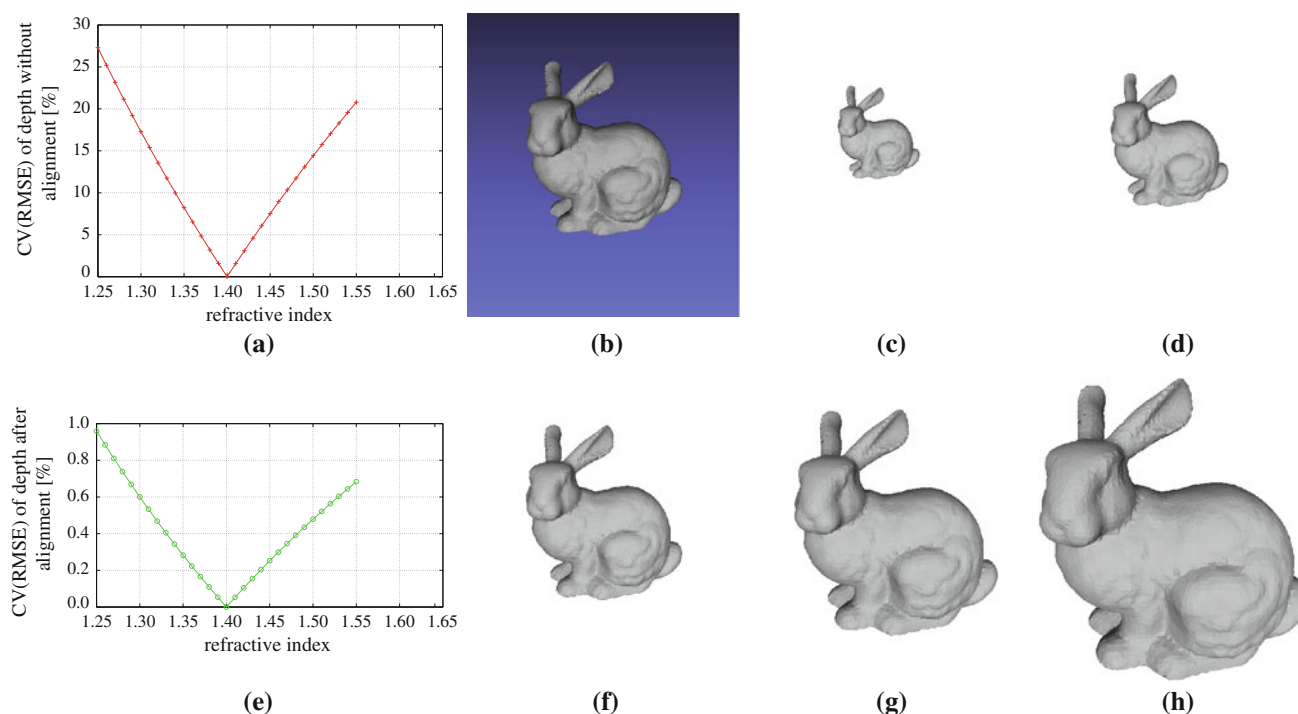
Fig. 12 Sensitivity of depth estimates with respect to the refractive index. (**a**) CV(RMSE) of depth without alignment. (**e**) CV(RMSE) of depth after alignment. (**b**) Ground truth bunny model. The rest of the figures show the reconstruction results using a refractive index of (**c**) 1.25, (**d**) 1.33, (**f**) 1.40, (**g**) 1.47 and (**h**) 1.55

three images. Figure 11a shows the RMSE of the estimates for the refractive index and CV(RMSE) of depth against different noise levels. The reconstruction result under a noise level of 1.0 pixel is shown in Fig. 11c, d.

In estimating the sensitivity of depth with respect to the refractive index, a synthetic experiment was carried out using varying refractive indices. Correspondences across a direct image and a refracted image of the bunny model were obtained, and used to estimate the depth. The ground truth refractive index was 1.4, and the refractive indices used were ranging from 1.25 to 1.55. CV(RMSE) of depth was measured and the reconstructed bunny models using five different refractive indices were shown in Fig. 12. The mean depths for different refractive indices were also computed. They differ from the mean depth for the ground truth refractive index by -13.9, -6.0, 0, 5.3, and 10.6 cm for a refractive index of 1.25, 1.33, 1.4, 1.47 and 1.55 respectively. This echoes the third observation in Sect. 4.1 that with all other parameters fixed, the estimated depth increases monotonically with the refractive index. Furthermore, with a larger refractive index, the reconstruction result has a larger scale, as shown in Fig. 12b–d, f–h. We also measured CV(RMSE) of depth after aligning the reconstructed bunny models with the ground truth (by a similarity transformation). CV(RMSE) of depth after alignment is shown in Fig. 12e. It is shown that while the CV(RMSE) of depth without alignment is

relatively large if the refractive index is not accurate (27 % for a refractive index of 1.25), CV(RMSE) of depth after alignment is still very small (less than 1 % for a refractive index of 1.25). It implies that the shape of the reconstruction result has little distortion even in the case of using an inaccurate refractive index. It can be seen in Fig. 12b–d, f–h that the reconstruction results using different refractive indices have almost the same shape as the ground truth. This fits human perception that when people see an object through a thick glass block, they can only feel little distortion. Also, note that the maximum difference of refractive indices within the visible spectrum for the glass used in the real experiments is only 0.018. If we consider the CV(RMSE) for the range $[1.4 - 0.018, 1.4 + 0.018]$, the error is still quite small even without the alignment.

6.2 Real Experiments

A common transparent glass (i.e., it is not of optical grade) with parallel planar faces and a thickness of 4.9 cm (measured by a common ruler) was used in the real experiments. Note that it follows from Eq. (10) that if the thickness of the medium is unknown, scene depths can still be obtained up to an unknown scale. In all of the following real experiments, the focal length of the camera was 24 mm, and the
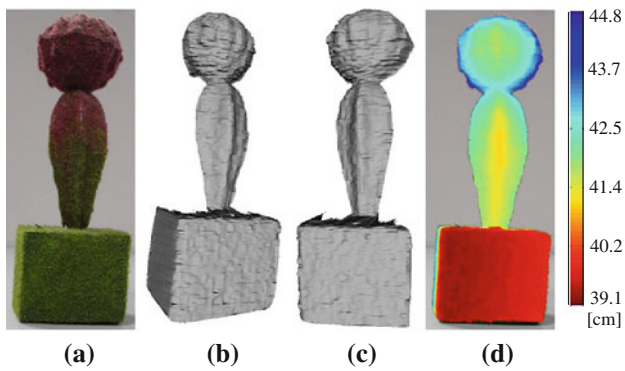
Fig. 13 Reconstruction of a flower model. (**a**) A direct image of the flower model. (**b**), (**c**) Two views of the reconstructed flower model. (**d**) A depth map of the flower model

intrinsic parameters of the camera were calibrated using Bouguet (2008).

In the first real experiment, one direct image and two refracted images of a flower model were captured. The flower model was segmented using Grady (2006) and had a resolution of 749×1823. The segmentation result was used to filter out correspondences outside the image of the flower model. Correspondences between the direct image and each of the two refracted images were first obtained using SIFT feature matching Lowe (2004), followed by a RANSAC Hartley and Zisserman (2004) procedure to eliminate outliers. In using RANSAC, two pairs of correspondences were sampled to obtain the focus of refraction as the intersection of their refraction lines, and inliers were identified by computing the distances between their refraction lines and the focus of refraction. This procedure was carried out to obtain correspondences between the direct image and the first refracted image as well as between the direct image and the second refracted image. The RMSE (in terms of pixel) of distance between the estimated vanishing point for the first refracted image and the ground truth is 362.09 pixels, and the RMSE is 305.10 pixels for the second refracted image. Compared to the distance of 5220.6 and 4078.9 pixels between the vanishing point and the image center for the first and second
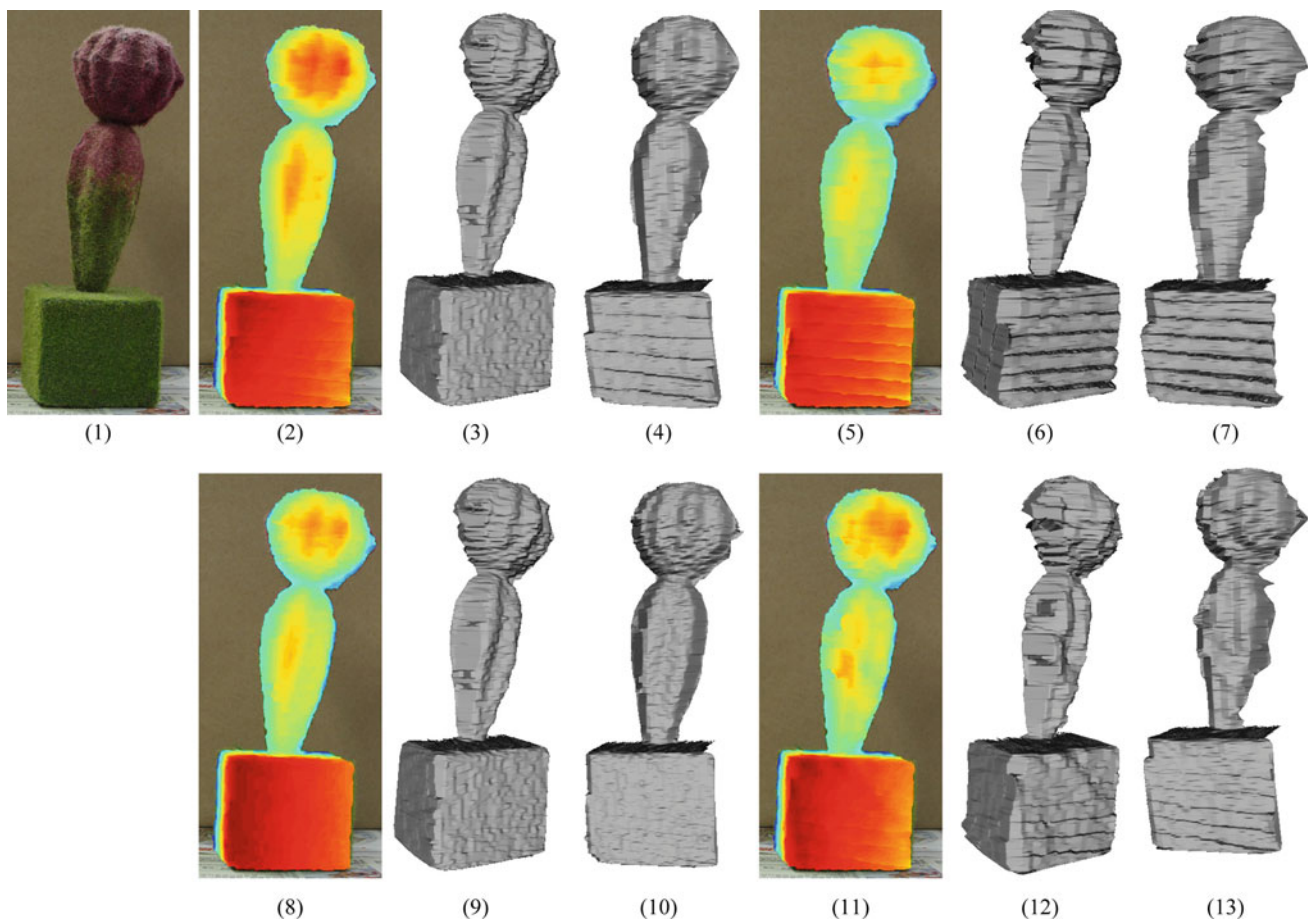

Fig. 14 Reconstruction of a flower model using media with different poses, different thickness and refractive indices. *1* A direct image of the flower model. Reconstruction results using *2–4* reference experiment, *5–7* smaller angle between the normal of the medium surface and the principal axis, *8–10* thicker medium, and *11–13* another medium with a higher refractive index
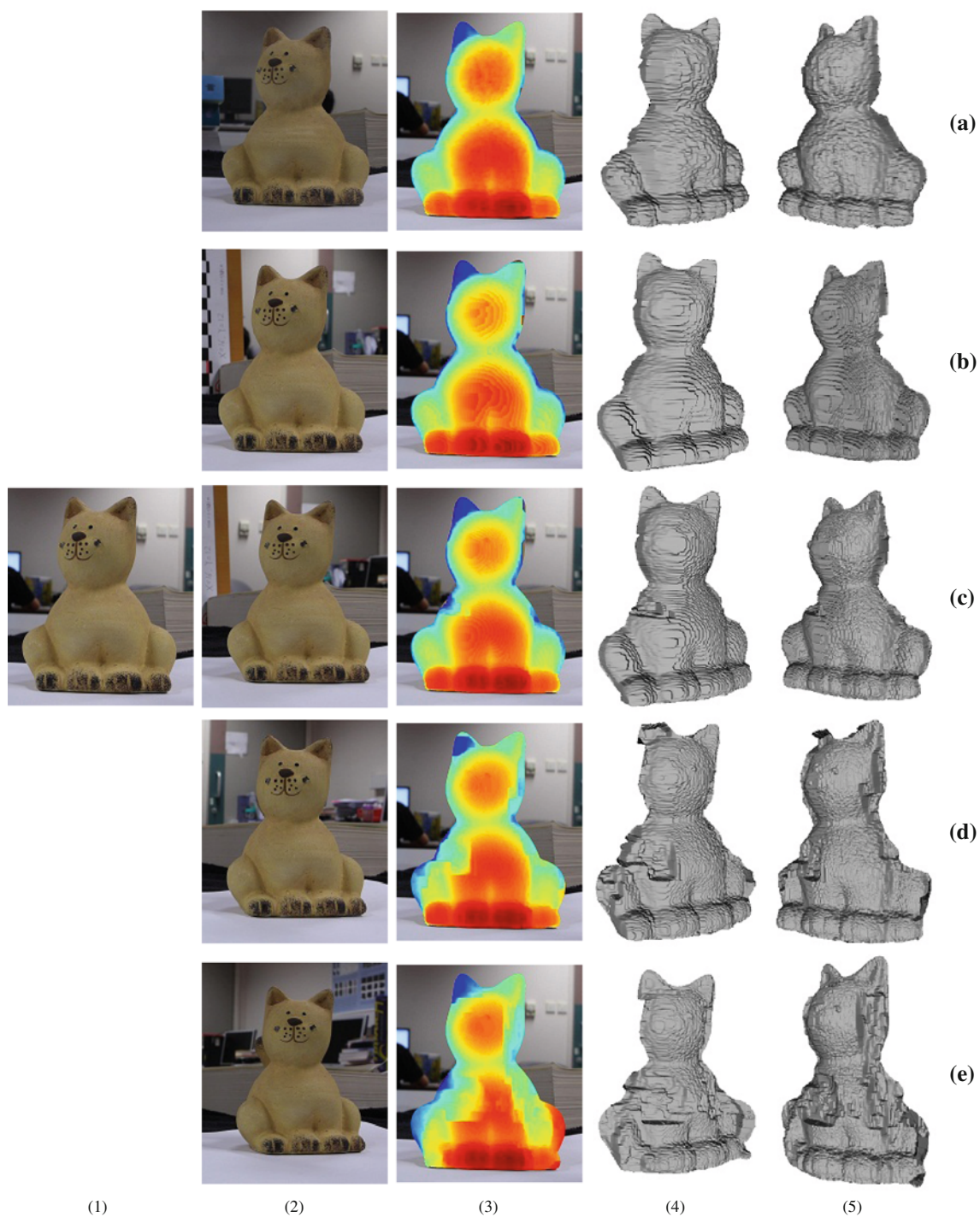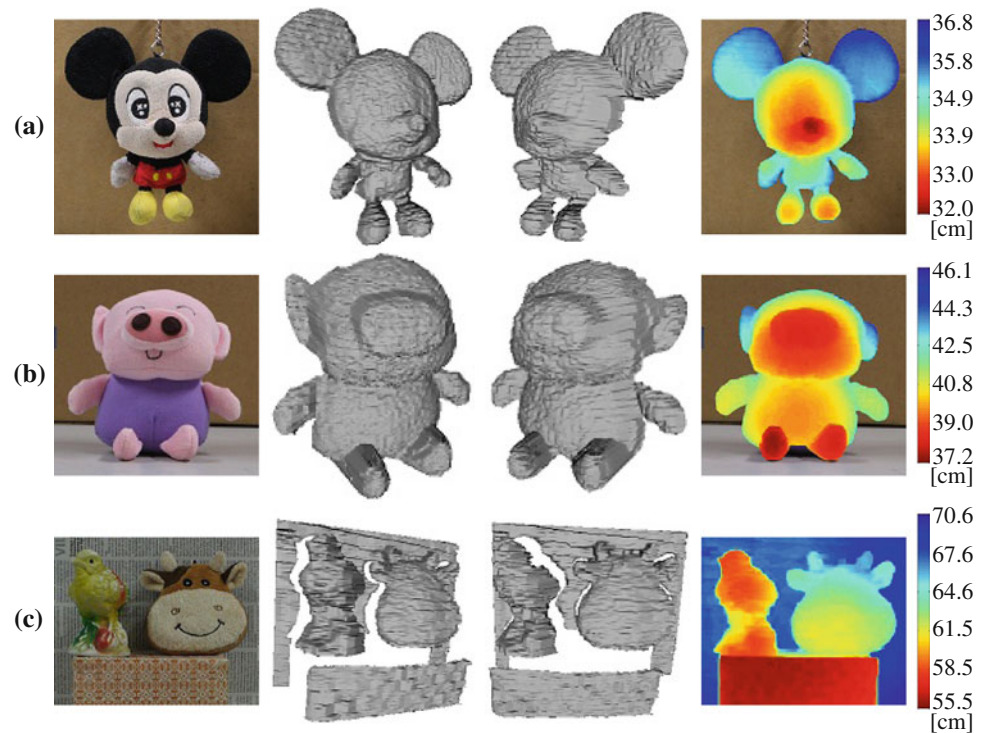
Fig. 15 Comparison between the proposed method and a standard stereo method. (**a**) Reconstruction using the proposed method. (**b**)–(**e**) Reconstructions using a standard stereo method with a baseline of 3.47, 5.95, 11.03 and 16.29 cm, respectively. The first column shows the first image used in both methods. The second column shows the refracted image used in the proposed method and the second images used in the stereo method. The third column shows the reconstructed depth maps of the cat model. The last two columns show two views of each reconstruction

refracted images respectively, the errors are small and the refraction line constraint still holds, which is proved by the fact that the RMSE of distance between correspondences in the first refracted image and the refraction lines is 0.80 pixel, and the RMSE of distance between correspondences in the second refracted image and the refraction lines is 0.74

Fig. 16 Reconstructions of (**a**) Mickey Mouse, (**b**) McDull, and (**c**) a scene consisting of two toy models. The *first column* shows the direct image of each scene; the *second* and the *third columns* show two views of each reconstruction; and the *fourth column* shows the recovered depth map of each scene

pixel. The correspondences across the three images were then established via points in the direct image, and were used to estimate the poses and the refractive index of the medium. SIFT flow (Liu et al. 2010) was next used to obtained dense correspondences between the direct image and each of the two refracted images, which were then used together with the estimated medium pose and refractive index to reconstruct the flower model (see Fig. 13). The experimental result was also quantitatively compared against the ground truth. In the experiment, the ground truth poses of the medium were obtained by taking an image of a checkerboard pattern placed on the medium surface. The angle between the estimated surface normal of the medium and the ground truth is 1.94° for the first pose and 2.17° for the second pose. The glass, made up of PMMA, has a refractive index range of [1.485, 1.503] within the visible spectrum. The estimated refractive index is 1.382. The reconstructed 3D points were also re-projected to the three images, and the RMSE of the re-projections is 1.178 pixels.

Real experiment was also carried out using media with different poses, thickness and refractive indices. In the reference experiment, the medium is made up of PMMA with a thickness of 3 cm, and it was positioned with an angle of 52.2° between the normal of the medium surface and the principal axis. The reconstruction result is shown in Fig. 14(2–4). Figure 14(5–7) show reconstruction result with the same setting as the reference experiment except that the angle between the normal of the medium surface and the principal axis was 38.3°.

Figure 14(8–10) show the reconstruction result with the same setting as the reference experiment except that the thickness of the medium was 4.9 cm. Figure 14(11–13) show the reconstruction result with the same setting as the reference experiment except that the medium is made up of crystal with a refractive index range of [1.539 1.560] within the visible spectrum. The RMSE of the re-projection errors of these four experiments were 1.880, 0.612, 2.056 and 1.867 pixels, respectively. It can be seen that reconstruction accuracy are not guaranteed to be improved using the three ways indicated in the synthetic experiments. For instance, using a thicker medium recovers a visually better result but with a larger re-projection error. The reason is that although the reconstruction result can be improved using the three ways shown in the synthetic experiments under the same noise level, in practice using different setting may cause different noise levels due to light absorbtion and reflection. In particular, thicker medium has more absorbtion and larger angle can cause more serious reflection effect. Therefore, reconstruction accuracy are not guaranteed to be improved using the three ways indicated in the synthetic experiments.

In the third real experiment, the proposed method was compared against a stereo method. One direct image and one refracted image were used to reconstruct a cat model. The orientation of the medium surface was estimated from correspondences in the two images obtained using SIFT feature matching. SIFT flow was then exploited to obtain dense correspondences across the two images, which were used to

reconstruct the cat model (see Fig. 15a). The reconstructed model was re-projected to the two images, and the RMSE of the re-projections is 0.6525 pixel. The cat model was also reconstructed using a standard stereo method using triangulation with a baseline of 3.47, 5.95, 11.03 and 16.29 cm, respectively. Dense correspondences of the cat model across the images were obtained similarly using SIFT flow, which were then used to reconstruct the cat model by triangulation (Hartley and Zisserman 2004). In order to better compare the two methods, the first image of the cat model used in both methods were identical. The experimental results with different baselines are shown in Fig. 15b–e. The re-projection errors using the stereo method with different baselines are 0.7592, 3.2381, 4.9769 and 7.3162 pixels, respectively. It is shown in Fig. 15 that the proposed method and the stereo method with a baseline of 3.47 cm achieved the best results. This can be explained by the fact that these two sets of images are similar and good dense correspondences can be obtained from SIFT flow. On the other hand, dense correspondences are difficult to achieve in the case when the two images suffered from large distortions due to translation, rotation and occlusions.

More experimental results of the proposed method are shown in Fig. 16. Figure 16a shows the reconstruction of Mickey Mouse, and the RMSE of the re-projections is 0.8102 pixel. Figure 16b shows the reconstruction of McDull (a famous Hong Kong cartoon character), and the RMSE of the re-projections is 0.8352 pixel. Figure 16c shows the reconstruction of a scene composed of two toy models, and the RMSE of the re-projections is 0.7520 pixel. Visually reasonable results were achieved in all these three experiments. Note that the artifacts in the reconstructions were mainly caused by inaccurate correspondences obtained using the SIFT flow method.

7 Conclusions

A novel method for depth from refraction is introduced in this paper. It is demonstrated that a transparent medium with parallel planar faces can be used to recover scene depth. Two images of a scene are captured by a camera with and without placing a transparent medium between the scene and the camera. Correspondences in images are then used to obtain the orientation of the parallel planar faces of the medium and the depths of scene points. It is further pointed out that a third image with the medium positioned in another orientation can be used to estimate the refractive index of the medium. Experiments on both synthetic and real data show promising results. With the proposed method, the pose and refractive index of the transparent medium, and depths of scene points can be estimated simultaneously. Nevertheless, the proposed method suffers from the same intrinsic limitation as other existing depth from refraction methods: it corresponds to a small baseline multiple-view approach. Hence the proposed method can mainly be used to recover depth for near scenes.

Appendix

Forward Projection for Refracted Images

In the special case where the parallel planar faces of the transparent medium are parallel to the image plane, the refracted projection of the 3D point $\mathbf{P}$ is obtained by directly projecting point $\mathbf{P}'$ to the image plane using the projection matrix (see Fig. 4). $\mathbf{P}'$ is the intersection of the line $\mathbf{PP}^*$ and the line passing through $\mathbf{O}$ and $\mathbf{S}_2$. Suppose the coordinates of $\mathbf{P}$ is (P_x, P_y, d) and the coordinates of $\mathbf{P}'$ is (P_x, P_y, d'). After applying simple trigonometrical transform to Eq. (6), we can obtain the following expression for $\tan \theta_2$:

$$\tan \theta_2 = \sqrt{\frac{\tan^2 \theta_1}{(n^2 - 1)\tan^2 \theta_1 + n^2}}. \tag{16}$$

Substituting Eq. (16) into Eq. (4) gives

$$\begin{aligned}
&(n^2 - 1)(d - w)^2 \tan^4 \theta_1 - 2d(d - w)(n^2 - 1)\tan \alpha \tan^3 \theta_1 \\
&+ [n^2(d - w)^2 - w^2 + (n^2 - 1)d^2 \tan^2 \alpha]\tan^2 \theta_1 \\
&- 2n^2 d(d - w)\tan \alpha \tan \theta_1 + n^2 d^2 \tan^2 \alpha = 0. \tag{17}
\end{aligned}$$

Solving Eq. (17) gives the solution to $\tan \theta_1$. Finally, d' can be expressed as

$$d' = \frac{d \tan \alpha}{\tan \theta_1}. \tag{18}$$

In the case where the parallel planar faces of the medium are not parallel to image plane, the refracted projection of the 3D point in the camera plane after rotation can be computed using the aforementioned method. A planar homography can then transform the refracted projection in the camera plane after rotation to the refracted projection in the original camera plane.

References

Agarwal, S., Mallick, S. P., Kriegman, D., Belongie, S. (2004). On refractive optical flow. In *Proceedings of European conference on computer vision* (pp. 483–494). Prague.

Bouguet, J. Y. (2008). Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/

Chen, Z., Wong, K. Y. K., Matsushita, Y., Zhu, X., Liu, M. (2011). Self-calibrating depth from refraction. In *Proceeding of international conference on computer vision* (pp. 635–642). Barcelona.

Gao, C., Ahuja, N. (2004). Single camera stereo using planar parallel plate. In *Proceedings of international conference on pattern recognition* (Vol. 4, pp. 108–111). New York.

Gao, C., Ahuja, N. (2006). A refractive camera for acquiring stereo and super-resolution images. In *Proceedings of conference computer vision and pattern recognition* (Vol. 2, pp. 2316–2323). New York.

Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligance*, *28*(11), 1768–1783.

Hartley, R. I., & Zisserman, A. (2004). *Multiple view geometry in computer vision*, (2nd ed.). Cambridge: Cambridge University Press.

Ishigure, T., Nihei, E., & Koike, Y. (1996). Optimum refractive-index profile of the graded-index polymer optical fiber, toward gigabit data links. *Applied Optics*, *35*, 2048–2053.

Lee, D., & Kweon, I. (2000). A novel stereo camera system by a biprism. *IEEE Transactions on Robotics and Automation*, *16*, 528–541.

Liu, C., Yuen, J., & Torralba, A. (2010). Sift flow: Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 978–994.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*, 91–110.

Maas, H. G. (1995). New developments in multimedia photogrammetry. In A. Gruen, & H. Kahmen (Eds.), *Optical 3-D measurement techniques III*. Karlsruhe: Wichmann Verlag.

Morris, N. J., & Kutulakos, K. N. (2011). Dynamic refraction stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 1518–1531.

Murase, H. (1992). Surface shape reconstruction of a nonrigid transport object using refraction and motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*, 1045–1052.

Nishimoto, Y., Shirai, Y. (1987). A feature-based stereo model using small disparities. In *Proceedings on IEEE international workshop on industrial applications of machine vision and machine intelligence (Seiken symposium)* (pp. 192–196). Tokyo.

Pentland, A. P. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*, 523–531.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, *47*, 7–42.

Sellmeier, W. (1871). Zur erklarung der abnormen farbenfolge im spectrum einiger substanzen. *Annalen der Physik und Chemie*, *219*, 272–282.

Shimizu, M., Okutomi, M. (2006). Reflection stereo-novel monocular stereo using a transparent plate. In *Proceedings on Canadian conference on computer and robot vision* (pp. 14 CD-ROM). Quebec.

Shimizu, M., Okutomi, M. (2007). Monocular range estimation through a double-sided half-mirror plate. In *Proceedings on Canadian conference on computer and robot vision* (pp. 347–354). Montreal.

Shimizu, M., Okutomi, M. (2008). Calibration and rectification for reflection stereo. In *Proceedings of conference on computer vision and pattern recognition* (pp. 1–8). Providence.

Subbarao, M., Gurumoorthy, N. (1988). Depth recovery from blurred edges. In *Proceedings of conference on computer vision and pattern recognition* (pp. 498–503). Puerto Rico.

Surya, G., Subbarao, M. (1993). Depth from defocus by changing camera aperture: A spatial domain approach. In *Proceedings of conference on computer vision and pattern recognition* (pp. 61–67). Vienna.

Zhang, X., & Cox, C. S. (1994). Measuring the two-dimensional structure of a wavy water surface optically: A surface gradient detector. *Experiments in Fluids*, *17*, 225–237.

Zhou, C., Cossairt, O., Nayar, S. (2010). Depth from diffusion. In *Proceeding of conference on computer vision and pattern recognition* (pp. 1110–1117). San Francisco.

Zhou, C., Lin, S., Nayar, S. (2009). Coded aperture pairs for depth from defocus. In *Proceedings of international conference on computer vision* (pp. 325–332). Kyoto.