Birefractive Stereo Imaging for Single-Shot Depth Acquisition

Seung-Hwan Baek[†]   Diego Gutierrez[*]   Min H. Kim[†]

[†]KAIST   [*]Universidad de Zaragoza, I3A

(a) Prototype   (b) Double refraction in calcite   (c) Captured through calcite in front of a camera   (d) Depth from double refraction
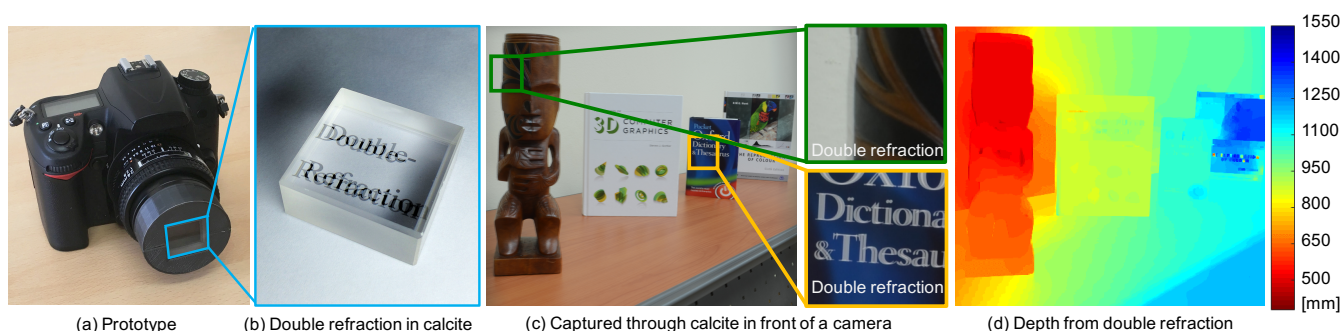
Figure 1: *(a) We present birefractive stereo for single-shot depth imaging, by simply placing a birefringent crystal in front of a camera. (b) Birefringent materials such as calcite produce a double-refraction phenomenon. (c) We rely on this effect and develop a novel birefringent imaging model to leverage the disparity information encoded in the captured images, obtaining (d) depth.*

Abstract

We propose a novel *birefractive* depth acquisition method, which allows for single-shot depth imaging by just placing a birefringent material in front of the lens. While most transmissive materials present a single refractive index per wavelength, birefringent crystals like calcite posses two, resulting in a double refraction effect. We develop an imaging model that leverages this phenomenon and the information contained in the ordinary and the extraordinary refracted rays, providing an effective formulation of the geometric relationship between scene depth and double refraction. To handle the inherent ambiguity of having two sources of information overlapped in a single image, we define and combine two different cost volume functions. We additionally present a novel calibration technique for birefringence, carefully analyze and validate our model, and demonstrate the usefulness of our approach with several image-editing applications.

Keywords: refractive stereo, depth, double refraction, birefringence

Concepts: •Computing methodologies → Computational photography; 3D imaging;

1 Introduction

Depth estimation from images is very useful for many applications besides computer graphics, like robotics, autonomous vehicles, or augmented reality. Common binocular/multiview stereo imaging techniques require at least two cameras to estimate disparity, resulting in additional cost and a large form factor. Many alternative approaches requiring a single camera have been proposed; however, they all present several shortcomings or require additional hardware, hindering their applicability in every-day scenarios. For instance, some methods rely on structured light [Zhang and Nayar 2006], while bi-prism stereo [Lee and Kweon 2000] reduces the effective sensor resolution by half; this trade-off between spatial resolution and the number of captured images is also one of the main limitations of light-field depth imaging. Reflection-based stereo [Yano et al. 2010] requires that the scene be imaged through a double-sided half-mirror plate, slanted at the right angle. Depth-from-defocus techniques [Levin et al. 2007; Bando et al. 2008] require a shallow depth-of-field to increase depth sensitivity, and can suffer from depth ambiguity due to low-frequency cues. Existing refraction-based stereo techniques [Gao and Ahuja 2006; Chen et al. 2013; Baek and Kim 2015; Baek and Kim 2016] are suitable for static scenes only, since they require multiple input images.

In this paper, we propose *birefractive* stereo, a novel technique for single-shot depth estimation from double refraction. This is easily achieved by simply placing a birefringent material in front of the lens of any conventional camera. It is therefore cheap, has a minimal impact on the form factor of the setup, does not sacrifice spatial resolution, introduces no visible image degradation (as our analysis and results show), and is readily useful for personal photography, allowing users to capture images without the burden of careful setups or complex additional hardware.

Birefringence is an optical property of some anisotropic, transmissive materials; an incident light ray is split into two rays (called ordinary and extraordinary), causing a double refraction [see Figure 1(b)]; we offer a more detail description in Section 2. This phenomenon allows us to capture a single image, containing overlapping information of the scene [shown in Figure 1(c)], from which depth information can be inferred [see Figure 1(d)].

However, there are many technical **challenges** that need to be overcome, to be able to estimate depth from double refraction: first, while some works have focused on birefringence for computer graphics simulations [Tannenbaum et al. 1994; Guy and Soler 2004; Weidlich and Wilkie 2008; Latorre et al. 2012], a description of the birefringence imaging model suitable for stereo imaging

does not exist. Although the geometric relationship between depth and pixel displacement through ordinary refraction has been well studied in traditional refractive stereo [Gao and Ahuja 2006], this model no longer works for birefractive stereo, given the particular characteristics of the extraordinary refraction. The relationship between depth and pixel displacement by double refraction has not been formalized yet. Moreover, existing refractive stereo calibration methods do not work in the case of birefringent materials. Last, traditional refractive stereo looks for correspondences in a pair of images (direct and refracted) [Baek and Kim 2016; Baek and Kim 2015]; however, since birefringence superimposes the two refracted images into a single one, a new correspondence search strategy must be devised, capable of handling overlapping information in the image.

To tackle these challenges, we introduce the following main **contributions**:

- We formulate a novel birefractive image formation model, introducing the *walk-off plane* to formally account for extraordinary refraction, which makes the mapping from extraordinary disparity to depth possible, enabling single-shot birefractive stereo.

- We introduce a novel edge-aware correspondence search algorithm in the gradient domain, which allows us to determine disparity in a superimposed double-refraction image.

- We propose a novel calibration method for birefractive stereo, which determines not only the physical orientation of the medium, but also the intrinsic orientation of its optical axis in screen coordinates.

- We show the advantages of our approach compared to other state-of-art methods, and demonstrate several applications including refocusing, anaglyph generation, automatic object segmentation, and depth-aware image editing.

Nonetheless, our system is not free from limitations. To capture double refraction across an image, a smaller aperture is preferable, which could lower the level of exposure. Our birefractive stereo utilizes gradients of double refraction. A lower ISO setting or additional denoising might be preferable. Despite this, we show that our birefractive stereo method acquires depth from a single-shot image in a wide range of scenarios, as demonstrated in the results.

2 Background

Bartholin discovered the phenomenon of *double refraction* in transmissive, *birefringent* crystals [Hecht 2002]. This is due to the microscopic structure of the medium (the ordered arrangement of the atoms in the lattice), and the different binding forces of molecules along axes. Figure 2 depicts this phenomenon: When the incoming ray $\mathbf{l}^i$ arrives at a plane surface, it splits into two rays, called ordinary (o-ray) $\mathbf{l}_o^r$ and extraordinary (e-ray) $\mathbf{l}_e^r$. These two rays become polarized in orthogonal directions, even in the case of an incoherent incident ray, a phenomenon that was first reported by Fresnel. The o-ray $\mathbf{l}_o^r$ travels on *the plane of incidence* through the medium, and its direction cosines are determined by Snell's law [Hecht 2002]; however, the e-ray $\mathbf{l}_e^r$ walks off[1] from the plane of incidence, and its direction cosines can no longer be determined by Snell's law. Instead, the direction of the e-ray is determined by the orientation of the birefringent crystal's optical axes, and can be calculated by applying Huygens's principle and the Maxwell equations [Avendaño-Alejo et al. 2002; Weidlich and Wilkie 2008; Latorre et al. 2012], or phase-matching modeling [Liang 1990].

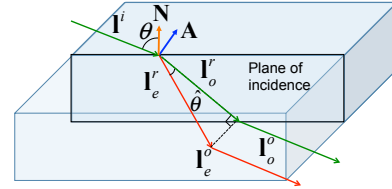[1]The *walk-off angle* is the separation between the o-ray and the e-ray in the crystal.

Figure 2: *Light transport in a birefringent crystal. When the incident ray $\mathbf{l}^i$ arrives at the surface, it splits into the ordinary (o-ray) $\mathbf{l}_o^r$ and the extraordinary (e-ray) $\mathbf{l}_e^r$ rays. The e-ray vector $\mathbf{l}_e^r$ walks off from the plane of incidence at angle $\hat{\theta}$, causing a double refraction.*

Most crystals are optically anisotropic, meaning that the propagation characteristics of a wave inside them depend on the direction of propagation. This is explained by the directional dependency of the dielectric permittivity tensor, which relates the electric field and the dielectric displacement vectors (this tensor becomes a scalar for isotropic media); its three orthogonal axes are called the *principal axes* of the crystal. If two of the three components are equal (the material has two different binding forces along the three axes), the crystal is uniaxial; otherwise, the crystal is biaxial (the material has three different binding forces along the three axes). The interested reader can refer to other sources in the optics field for an in-depth analysis of the phenomenon [Hecht 2002], or a more detailed explanation adapted to computer graphics [Weidlich and Wilkie 2008; Latorre et al. 2012].

3 Related Work

Simulating Birefringence A few works in computer graphics have dealt with the simulation of birefringent materials. Tannenbaum et al. [1994] proposed a matrix-based formulation for uniaxial crystals, deriving formulas for the propagation of the extraordinary ray. A real-time implementation was presented by Guy and Soler [2004], based on careful approximations focused on the rendering of gems. Weidlich and Wilkie [2008] integrate uniaxial birefringence into a conventional ray tracer, and derive the complete set of formulas needed to simulate the effect in a physically accurate way. Later, Latorre et al. [2012] present the first solution to deal with biaxial crystals, relying on numerical approximations.

Capture and Display Some practical imaging applications of birefringent crystals have been proposed in the literature. Zalevsky and Ben-Yaish [2007] extended the depth-of-field (DOF) of a camera using a birefringent material carefully placed between the lens and the sensor, while Tsai and Brady [2013] introduced a hyperspectral imager that makes use of two birefringent slabs, along with a coded aperture. On the display side, Shestak et al. [2015] recently devised a stereo television system which requires glasses made of birefringent materials. To the best of our knowledge, ours is the first method that leverages birefringence to estimate depth from a single shot.

Reflective Stereo The acquisition setup for reflective stereo requires that the scene being captured be observed through a double-sided half-mirror plate, positioned at a slanted angle [Shimizu and Okutomi 2006; Yano et al. 2010]. However, imaging both the front and rear reflections requires a large angle between the mirror's normal and the camera's view direction. This results in a large form factor, and a careful scene setup, which is impractical for daily use; our method overcomes these limitations.

Refractive Stereo Refractive stereo estimates depth from displacement of either a pair of direct/refracted images or a pair of refracted images using different media [Chen et al. 2013; Gao and Ahuja 2006; Baek and Kim 2015; Baek and Kim 2016]. Since

the two images cannot be captured in a single shot, the method is limited to static scenes. On the contrary, our birefractive stereo technique takes into account the displacement in a superimposed image caused by the refracted o-ray and e-ray on the same image, removing the need for multiple shots.

Single-Shot Depth Single-shot depth imaging allows to capture depth from scenes. Several applications have been proposed in the fields of computer vision and robotics. Lee and Kweon [2000] introduced a binocular stereo method; they use a bi-prism to introduce disparity between two images, splitting the camera imaging space in two. Unfortunately, the image resolution is reduced by half accordingly. In addition, spectral dispersion by the prism blurs images severely, hindering correspondence search. This degrades depth accuracy since feature points cannot be correctly identified [Li and Wang 2009]. Levin et al. [2007] introduced a principled coded aperture approach to infer depth from defocus, while Bando et al. [2008] later proposed a color-coded aperture to produce different blur kernels or displacements in the three color channels. Subsequent testing of cross correlations across color channels allows for depth estimation. These aperture-based methods require a large aperture to achieve a shallow depth of field, in order to improve depth sensitivity, and the approach can also suffer from depth ambiguity due to low-frequency cues of blurriness. Shi et al. [2015] obtain depth from small-scale defocus blur, an information that can be weakened if the image is downsampled. Also, the method cannot resolve edge smoothness ambiguity: a blurred edge may come from a latent sharp edge affected by defocus blur, or a smoother edge with potentially no blur. Different from these methods, our birefractive stereo technique does not require any camera modifications, nor a particular scene setup. Instead of analyzing edge blur, it leverages the information from a double-refraction image produced by simply placing a flat, birefractive optical filter in front of any conventional camera, avoiding the problem of edge smoothness ambiguity.

4 Overview

Different from ordinary refractive stereo, our goal is to leverage the displacement between the o-ray and the e-ray, to estimate depth z. This will allow us to enable *singlet-shot refractive stereo*, without needing to capture an additional direct image. Our birefractive stereo method consists of: (1) a formal description of our image formation model, which allows us to establish the relationship between o- and e-ray disparity, and depth; (2) a method to obtain depth information from a single image, where ordinary and extraordinary information appear superimposed, based on our gradient-domain search and dual derivation of matching cost volumes; and (3) a novel calibration method for the birefractive material and our imaging setup, to obtain the parameters of our correspondence model. These steps will be described in the three following sections.

5 Birefractive Image Formation Model

Figure 3 presents a schematic view of our birefractive image formation model, based on the background introduced in Section 2. Table 1 describes all the symbols and notation used in this section, while Table 2 enumerates each plane and its vectors, to provide a quick reference for the readers. In the traditional (non-refractive) model, a scene point P_s is projected to point P_d on the sensor through the optical center O of the lens (black line). When a refractive medium is placed in front of the lens, P_s will instead project through O to a new point P_o, according to Snell's law (green line). We denote as P_o^1 and P_o^2 the first and second intersections with the surfaces of the refractive medium, respectively. Since vectors $\overline{P_sP_o^1}$, $\overline{P_o^1P_o^2}$, and $\overline{P_o^2P_o}$ lie on the plane of incidence, the entire o-ray transport remains on this plane of incidence.

	Symb.	Description
Common Variables	O	Optical center (pin-hole) of a camera
	f	Focal length of a lens
	η_o	o-ray refractive index of the biref. medium
	η_e	e-ray refractive index of the biref. medium
	$\mathbf{a}$	Optical axis of the biref. medium
	$\mathbf{n}$	Normal vector of the plane surface of the biref. medium
	t	Thickness of the biref. medium
	E	Essential point defined by the biref. medium's normal
	E'	Back-projected essential point to the scene plane
	P_s	3D scene point
	P_d	Projected pixel position without the biref. medium
o-ray (Section 5.1)	P_o	o-ray pixel position onto the image plane
	P_o'	Back-projected point of P_o onto the scene plane
	P_o^1	Point where o-ray is at the rear face of the biref. medium
	P_o^2	Point where o-ray is at the front face of the biref. medium
	r_o'	Distance between P_s and P_o'
	r_o	o-disparity; Distance between P_d and P_o
	d_o	Walk-off length of o-ray
	θ_o^p	Angle between $\overline{EP_o}$ and $\overline{P_oO}$
	θ_o^i	Angle between $\overline{P_oO}$ and $\overline{EO}$
e-ray (Section 5.2)	P_e	e-ray pixel position onto the image plane
	P_e'	Back-projected point of P_e onto the scene plane
	P_e^1	Point where e-ray is at the rear face of the biref. medium
	P_e^2	Point where e-ray is at the front face of the biref. medium
	r_e'	Distance between P_s and P_e'
	r_e	e-disparity; Distance between P_d and P_e
	d_e	Walk-off length of e-ray
	θ_e^p	Angle between $\overline{P_dP_e}$ and $\overline{P_eO}$
	θ_e^i	Angle between $\overline{P_eO}$ and $\overline{EO}$
	θ_e^d	Angle between $\overline{P_eE}$ and $\overline{P_eP_d}$
	θ_e^m	Angle between $\overline{P_e^1P_e^2}$ and $\overline{P_e^2P_e^3}$

Table 1: *Symbols and notation used in the paper.*

Plane	Vectors belonging to the plane
Plane of incidence	$\overline{P_sP_o^1}$, $\overline{P_o^1P_o^2}$, $\overline{P_o^2P_o}$, $\overline{P_oP_d}$, $\overline{P_o'P_s}$, $\overline{P_oE}$, $\overline{P_o'O}$, $\overline{P_oO}$, $\overline{P_oE}$, $\overline{OE}$, $\hat{\mathbf{d}}_o$
Walk-off plane	$\overline{P_sP_e^1}$, $\overline{P_e^1P_e^2}$, $\overline{P_e^2P_e}$, $\overline{P_eO}$, $\overline{P_eP_d}$, $\overline{P_e'P}$, $\overline{P_e'O}$, $\hat{\mathbf{d}}_\mathbf{e}$
Scene plane	$\overline{P_o'P}$, $\overline{P_e'P_s}$
Sensor plane	$\overline{P_eP_d}$, $\overline{P_oP_d}$, $\overline{P_oE}$, $\overline{P_eE}$, $\hat{\mathbf{d}}_o$, $\hat{\mathbf{d}}_e$, $\mathbf{n}_e^p$, $\mathbf{e}_e^p$

Table 2: *Planes and vectors in our image formation model.*

If we place instead a *birefringent* medium, two rays from P_s will pass through O and reach the sensor at different positions: the ordinary ray at P_o (same as before, since it also follows Snell's law), and the extraordinary ray at P_e (purple line). Note that our imaging model differs from other birefringence studies [Weidlich and Wilkie 2008; Latorre et al. 2012] in that we concern ourselves with the refracted rays that go through O. Thus our model includes two different incoming rays $\overline{P_sP_o^1}$ and $\overline{P_sP_e^1}$, instead of one single incident ray from which the o-ray and the e-ray are produced. This allows us to define the *walk-off plane*, where vectors $\overline{P_sP_e^1}$, $\overline{P_e^1P_e^2}$, and $\overline{P_e^2P_e}$ lie; this is a key aspect of our model, since the entire e-ray transport now remains on this plane; the walk-off plane in extraordinary refraction is analogous in this regard to the plane of incidence for ordinary refraction.

Another important feature in our birefractive image formation model is the *essential point E* on the sensor plane, defined as the point where the vector $\mathbf{n}$ normal to the surface of the refracting medium intersects the sensor plane via the optical center O; note that $\mathbf{n}$ does not necessarily coincide with the optical axis of the lens in our method; therefore the essential point E is mainly determined by the orientation of the transmissive medium (this point will be obtained through calibration in Section 7). Points P_d, P_o and E lie on the so-called *essential line*, while P_e does not. This property will later allow us to narrow down the search range of correspondences.
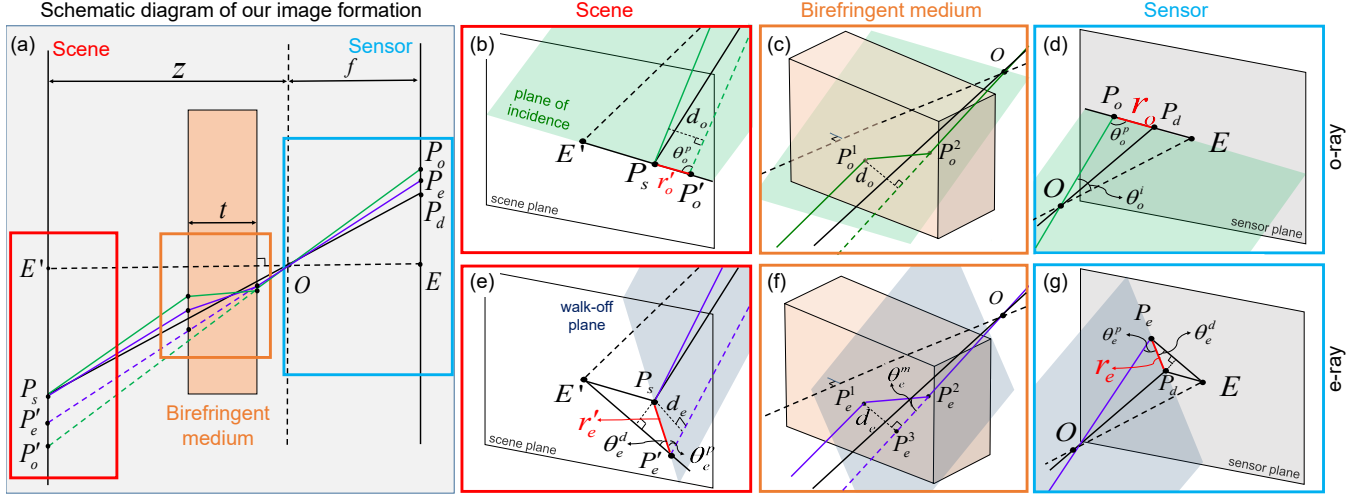
Figure 3: *(a) Cross-section schematic diagram of our image formation model with double refraction (see the text for the full details). In the absence of a refractive medium, scene point P_s is directly projected to P_d on the sensor plane. Introducing a birefringent medium, stereo rays from P_s are projected to P_o for o-ray and P_e for e-ray on the sensor. The first row $(b)-(d)$ shows close-up diagrams for o-ray light transport at the scene plane, the medium and the sensor plane, respectively. Note that the* entire *o-ray transport remains on the* plane of incidence *(Section 5.1). The second row $(e)-(g)$ presents similar diagrams for e-ray light transport (Section 5.2). By defining the* walk-off plane*, we ensure that the entire e-ray transport remains on such plane, which makes the mapping from e-ray disparity to depth possible.*

We seek to define two functions relating the captured ordinary and extraordinary ray intersections on the sensor, P_o and P_e, with depth: $\psi_{o\to e}(P_o, z) = P_e$, and $\psi_{e\to o}(P_e, z) = P_o$. For that, we first need to derive separate expressions for the o-ray (e-ray), that describe the relationship between P_o (P_e), P_d, and z. These expressions will then be combined to remove the P_d dependency (since our single-shot approach does not require a direct capture without the birefringent medium, so P_d is never imaged), and used to estimate depth in Section 6.

For the sake of completeness, we first briefly introduce how to infer depth from the disparity of the ordinary ray, a well-known technique that allows for two-shot, depth estimation methods [Gao and Ahuja 2006; Chen et al. 2013]. We will then develop depth from the extraordinary refraction, and show how to combine the two different sources of information present in a double-refraction image. Dependencies on known parameters such as the optical axis $\mathbf{a}$, normal orientation $\mathbf{n}$, and the refractive indices for the o-ray η_o, and the e-ray η_e will be solved after calibration (Section 7).

5.1 Ordinary Ray Disparity to Depth

As we have seen, scene point P_s projects to P_o through O in the presence of a refractive medium. The distance $r_o = |\overline{P_oP_d}|$ on the sensor plane is called the (ordinary) *refractive disparity*, which depends on the thickness t, and the index of refraction η_o of the medium. Let P_o' be a virtual projection of P_o through O on the scene plane, assuming no refraction. Given the focal length f, we can estimate the depth z of point P_s as [Chen et al. 2013]:

$$z = (r_o'/r_o)\,f, \tag{1}$$

where $r_o' = |\overline{P_o'P_s}|$ is defined on the scene plane. Note that since we deal with ordinary refraction, both r_o and r_o' lie on the plane of incidence. From basic trigonometry we have $r_o' = d_o/\sin\theta_o^p$, where $\theta_o^p = \angle(\overline{P_o'O}, \overline{P_o'P_s}) = \angle(\overline{P_oO}, \overline{P_oE})$. d_o is called the *lateral displacement* of the o-ray [Hecht 2002], and is computed as the distance between the two parallel vectors $\overline{P_oP_o^2}$ and $\overline{P_o^1P_s}$:

$$d_o = \left(1 - \sqrt{\frac{1-\sin^2\theta_o^i}{\eta_o^2 - \sin^2\theta_o^i}}\right) t\sin\theta_o^i\,,$$

where $\theta_o^i = \angle(\overline{OE}, \overline{OP_o})$. Equation (1) defines r_o as a function of depth z. We can now define a function $\psi_{o\to d}$, which provides a mapping between the captured o-ray position P_o and r_o, and the direct ray position P_d, as

$$\psi_{o\to d}(P_o, z) \equiv P_o + r_o(z)\,\hat{\mathbf{d}}_o = P_d, \tag{2}$$

where $\hat{\mathbf{d}}_o$ is the unit vector corresponding to $\overline{P_oE}$, defined on the essential line.

Different from traditional depth from refractive stereo algorithms, we do not know P_d, since no direct image is taken in our approach. In the following, we will derive a similar function $\psi_{e\to d}$ for the extraordinary ray. Combining both functions $\psi_{o\to d}$ and $\psi_{e\to d}$ will allow us to bypass the need for a second (direct) capture, enabling single-shot birefractive stereo.

5.2 Extraordinary Ray Disparity to Depth

While the o-ray follows Snell's law, the e-ray walks off from the plane of incidence following Huygens's principle and Maxwell's equations. A key concept in our derivation is therefore the *walk-off plane*.

Walk-off Plane The walk-off plane is defined by the incident ray $\overline{P_sP_e^1}$, and the outgoing e-ray $\overline{P_e^2P_e}$ through O so that the entire e-ray transport remains on the plane. Also belonging to that plane, we find vectors $\overline{P_eP_d}$ on the sensor plane, and $\overline{P_eO}$. Since we do not have a direct capture to obtain P_d, we use $\overline{P_e^2P_e^1}$ instead to obtain the normal $\mathbf{n}_e$ of the walk-off plane; given $\overline{P_eO}$, the direction cosines of the e-ray defining $\overline{P_e^2P_e^1}$ can be obtained by phase matching, using Liang's method [1990] (described in detail in the supplemental material).

Depth from E-ray Disparity Similar to the o-ray disparity to depth derivation, we define $r_e = |\overline{P_eP_d}|$ on the sensor plane, as the *extraordinary ray disparity*, which depends on the medium's thickness t, and its index of extraordinary refraction η_e. Let P_e' be a virtual projection of P_e through O on the scene plane, assuming no refraction; from the two triangles $\triangle(P_sOP_e')$ and $\triangle(P_dOP_e)$ on

the walk-off plane, we can define a relationship between depth z and e-ray disparity r_e as follows:

$$z = (r'_e / r_e) f, \qquad (3)$$

where $r'_e = |\overline{P'_e P_s}|$ is defined on the scene plane, and both r_e and r'_e lie on the walk-off plane. Trigonometric relations give us $r'_e = d_e / \sin \theta^p_e$, where $\theta^p_e = \angle(\overline{P'_e O}, \overline{P'_e P_s}) = \angle(\overline{P_e O}, \overline{P_e P_d})$. We call d_e the *lateral displacement* of the e-ray, and is computed as the distance between the two parallel vectors $\overline{P_e P_e^2}$ and $\overline{P'_e P_s}$. To obtain d_e, we define a triangle $\triangle(P_e^1 P_e^2 P_e^3)$ [see Figure 3(f)]; the angle $\theta^m_e = \angle(\overline{P_e^1 P_e^2}, \overline{P_e^2 P_e^3})$ can be obtained from the normalized dot product of $\overline{P_e^2 P_e^1}$ and $\overline{P_e O}$, then $d_e = |\overline{P_e^2 P_e^1}| \sin \theta^m_e$. The angle θ^p_e can be obtained from the normalized dot product of $\overline{P_e O}$ and $\overline{P_e P_d}$ from two triangles $\triangle(P_s O P'_e)$ and $\triangle(P_d O P_e)$ on the walk-off plane [see Figures 3(e) and (g)].

From Equation (3), we can define a function $\psi_{e \to d}$ mapping the captured e-ray position P_e and r_e, and the direct ray position P_d as

$$\psi_{e \to d}(P_e, z) \equiv P_e + r_e(z) \hat{\mathbf{d}}_\mathbf{e} = P_d, \qquad (4)$$

where $\hat{\mathbf{d}}_\mathbf{e}$ is the unit vector corresponding to $\overline{P_e P_d}$, defined on the intersection line between the walk-off and the sensor plane. $\hat{\mathbf{d}}_\mathbf{e}$ defines the direction of e-ray disparity. However, different from the o-ray derivation, we cannot rely on the essential point E to define $\hat{\mathbf{d}}_\mathbf{e}$, since P_e, P_d, and E are not co-aligned.
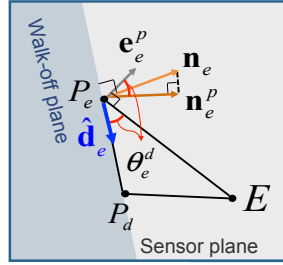

Figure 4: *Calculation of $\hat{\mathbf{d}}_\mathbf{e}$ [closeup view of Figure 3(g)].*

Instead, $\hat{\mathbf{d}}_\mathbf{e}$ can be obtained from the geometric relationship between the walk-off plane's normal $\mathbf{n}_e$ and $\overline{P_e E}$ (see Figure 4). First, we project $\mathbf{n}_e$ onto the sensor plane, obtaining $\mathbf{n}^p_e$. We then define a unit vector $\mathbf{e}^p_e$, perpendicular to $\overline{P_e E}$ on the sensor plane. From the normalized dot product of $\mathbf{n}^p_e$ and $\mathbf{e}^p_e$, we obtain $\theta^d_e = \angle(\overline{P_e P_d}, \overline{P_e E})$. This angle allows us to determine the unit vector $\hat{\mathbf{d}}_\mathbf{e}$ corresponding to $\overline{P_e P_d}$, by rotating the normalized vector of $\overline{P_e E}$ with θ^d_e-degrees around point P_e on the sensor plane. The sign of $\mathbf{n}^p_e \times \overline{P_e E}$ determines the direction of rotation, i.e., the positive sign indicates clockwise rotation, and vice versa.

5.3 Combining the two functions

Up to this point, we have obtained two analytic mapping functions [Equations (2) and (4)] defining the relationship between depth z, and o- and e-ray disparities, respectively. However, they both depend on P_d, which is not captured with our single-shot approach. To get rid of this dependency, we first invert these two functions to yield: $\psi_{d \to o}(P_d, z) = P_o$ and $\psi_{d \to e}(P_d, z) = P_e$. These mapping functions include a combination of multiple sinusoidal functions defining the o-/e-ray relationship, which makes them expensive to calculate analytically. Therefore, we instead pre-compute two tables storing values for P_o and P_e respectively, for each pair of input parameters P_d and z. Since the two functions now depend on the same input parameters, we combine the two tables into one, describing the four-dimensional mapping function $\psi_{d \to o,e}(P_d, z) = (P_o, P_e)$, from which the following two reciprocal functions can be obtained:

$$\begin{cases} \psi_{o \to e}(P_o, z) = P_e, \\ \psi_{e \to o}(P_e, z) = P_o. \end{cases} \qquad (5)$$

This correspondence model will be later used to calculate o- and e-ray disparity for each depth.

6 Depth from a Double-Refraction Image

Existing depth-from-stereo algorithms start with a set of potential z candidates, and the goal is to assign a depth value to each pixel P, given its left (x, y) and right (x', y') views. The relationship between corresponding pixel coordinates in both views is given by the epipolar line equation $x' = x + r$ (with $y' = y$), where r indicates disparity, which is inversely proportional to depth d. A *cost volume function* $C(P, z)$ is defined, which indicates how well two corresponding pixels match, using this equation; in other words, it stores the cost of assigning each depth candidate to each pixel. To improve performance, an edge-aware filtering process, so-called *cost aggregation*, is required for every depth slice in C.

Different from traditional stereo algorithms, there are two main challenges when estimating correspondences in double-refraction images. First, since both the ordinary and the extraordinary images are combined in a single one, corresponding pixels may have different colors due to overlapping [see Figure 5(a)]; this means that typically used correspondence algorithms based on color similarity cannot be applied. Instead, we propose a sparse correspondence search in the gradient domain. The second problem has to do with the inherent ambiguity in the double refraction, which makes it difficult to tell the ordinary from the extraordinary images. We overcome this by defining two different per-pixel, cost volume functions for double refraction C_o and C_e, and searching through the depth candidates using the mapping functions $\psi_{o \to e}$ and $\psi_{e \to o}$ [Equation (5)]. We then propagate our sparse estimated depths to the complete image. This is described in the following paragraphs.

Gradient-Domain Search We capture a superimposed image Y, which is the sum of two stereo images from the o- and e-ray, X_o and X_e. Note that one image is the displaced version of the other. We first calculate the gradient vectors $\partial \mathbf{Y}$ along the x and y directions[2], and define a map Φ made of all pixels P with a gradient magnitude $|\partial \mathbf{Y}(P)|_2 \geqslant \epsilon$, where $||_2$ represents the $l2$-norm. Estimating depth on these pixels by selectively using distinct gradients is helpful to mitigate the effects of noise. We empirically set $\epsilon = 0.01$ for all the results shown in this paper. Only these detected pixels $P \in \Phi$ have a correspondence cue on Y; we therefore estimate depth only on Φ, and propagate the results to every pixel in Y later.

The detected gradients $\partial \mathbf{Y}(P)$ on an edge pixel could belong to either X_o or X_e; when P belongs to X_o, its corresponding edge pixel should belong to X_e, and vice versa. However, as shown in Figure 5(b), these gradients may overlap when the orientation of an edge coincides with the direction of disparity in the double-refraction image, making it impossible to assign to X_o or X_e. To make our depth estimation robust, we exclude from Φ all pixels, whose gradient direction is close to the direction of disparity. We take a conservative approach and set a threshold angle of ± 45 degrees [see Figure 5(c)]. Moreover, the gradient magnitudes of corresponding pixels from double refraction are virtually the same (except at grazing angles on metallic surfaces, as discussed in Section 11); this is illustrated in Figures 6(a) and (b). This observation enables us to test whether two given pixels form a valid correspondence pair by comparing their gradient magnitudes. Our matching cost functions (described in the following paragraphs) will compare potential stereo correspondences in this gradient domain, rather

---

[2] To handle noise, we employ the kernel operator $\frac{1}{8}[-1, -2, -1; 0, 0, 0; 1, 2, 1]$ for x-gradients [Barron and Malik 2013], and its transposed version for y-gradients.
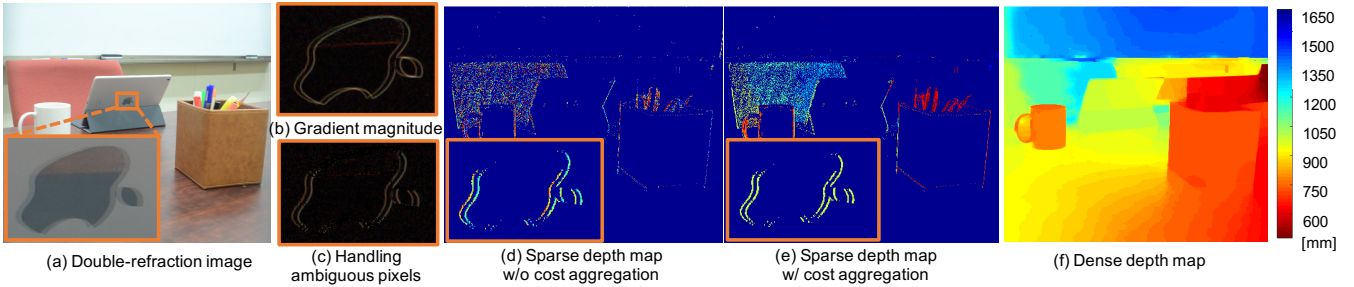
Figure 5: *Overall pipeline of our depth estimation. (a) Captured double-refraction image, which contains overlapped disparity information. This makes searching for correspondences a challenging task. (b) Gradient map reveals the positions of corresponding points more clearly, although there is some ambiguity where gradient directions coincide with disparity directions. (c) Ambiguous gradients are removed. (d) Estimated sparse depth without cost aggregation. (e) Estimated sparse depth with cost aggregation. (f) A propagated dense depth map.*

Figure 6: *Schematic diagrams of (a) two separate images, and (b) an overlapped image of double refraction. (a) The o-ray (blue line) and e-ray images (yellow line) have approximately the same per-pixel intensity (except at grazing angles on metallic surfaces), with one image being the translated version of the other. (b) The overlapped images consequently share the same gradient for corresponding pixels.*

than in color space, leveraging the special characteristics of double refraction.

Dual Matching Cost Volumes At this point, we still cannot disambiguate whether an edge pixel in Φ belongs to the ordinary X_o, or to the extraordinary refraction image X_e. We therefore calculate two matching cost functions for all gradient pixels, considering both options, as:

$$C_o(P,z) = |\partial \mathbf{Y}(P) - \partial \mathbf{Y}(\psi_{o\to e}(P,z))|_1,$$
$$C_e(P,z) = |\partial \mathbf{Y}(P) - \partial \mathbf{Y}(\psi_{e\to o}(P,z))|_1.$$

Cost function $C_o(P,z)$ evaluates the matching costs of pixel P on the o-ray image X_o for each depth z, finding its corresponding e-ray disparity using $\psi_{o\to e}$; on the other hand, $C_e(P,z)$ evaluates the matching costs of P on X_e at each z, using $\psi_{e\to o}$. To obtain these matching costs, we calculate the $l1$-norm of the difference between the two gradient vectors. We then disambiguate the nature of the refraction by selecting the cost function that yields the smallest error for all depth candidates $z_\forall$:

$$C(P,z) = \begin{cases} C_o(P,z), & \text{if } \min C_o(P,z_\forall) \leqslant \min C_e(P,z_\forall) \\ C_e(P,z), & \text{otherwise.} \end{cases}$$

We calculate this cost function for the red, green and blue channels separately, and add them to obtain the total cost. Since the matching cost function is a per-pixel operator, estimated matching costs in an image tend to be sparse and noisy. We therefore apply an edge-aware filter [Yang 2012] to each depth level in the cost volume C (this is called *cost aggregation* in traditional stereo), to achieve robust performance [see Figures 5(d) and (e)]. We finally select the most plausible depth estimate Z for each pixel P as:

$$Z(P) = \arg\min_z C(P,z),$$

and propagate these values to the rest of the image using Matting Laplacian optimization [Levin et al. 2004] [see Figure 5(f)].

7 Calibration of Birefringence

Since no established calibration method for birefractive stereo imaging exists, we present here a simple but practical method. To determine the intrinsic properties of the camera, such as its projection matrix and lens distortion coefficients, we can choose any existing camera calibration method; in this work we follow the checkerboard-based method of Zhang [2000]. In the following, we describe how to calibrate the optical properties of the birefringent material, in particular its essential point, as well as its optical axis. Note that this calibration process only needs to be performed once.

Essential Point In our method, the normal $\mathbf{n}$ of the birefractive medium can be oriented in any direction. Determining the orientation of $\mathbf{n}$ is equivalent to determining the location of E on the sensor plane. To do so, we first take three different images of a checkerboard as input: a direct image without the medium, plus two linearly-polarized images with the medium, to image the o-ray (point P_o on the sensor) and the e-ray (P_e), respectively [see Figure 7(a)]. To correctly identify ordinary from extraordinary points, we follow traditional refractive stereo techniques, and leverage the fact that the essential lines of o-rays should converge to E [Chen et al. 2013] on the sensor plane; e-rays, on the contrary, do not converge in general except for very particular geometric relationships. We therefore superimpose a pair of our checkerboard images (direct, plus each one with the refractive medium, separately) and connect corresponding points to elucidate which image contains the ordinary or the extraordinary refractions. In particular, in order to estimate the coordinates of the essential point E, we define the N connecting lines (N=77 in our implementation) as implicit functions $l_i = \{(x,y)|\, a_i x + b_i y + c_i = 0\}$, and solve a linear system with N equations using least-squares approximation. We measure line convergence by evaluating the total variation of the essential-line convergence, calculating the error of the optimization as $\sum_{i\in\{1...N\}} |a_i E_x + b_i E_y - c_i|$, where $E = [E_x, E_y]^\top$ lies on the sensor plane. The smaller error (about half in our experiments) indicates the o-ray image, which in turn yields the coordinates of E.

Optical Axis Commercial off-the-shelf uniaxial crystals are calibrated, and their optical axis of birefringence is provided. However, for birefractive stereo, this optical axis needs to be calibrated with respect to the camera coordinates.
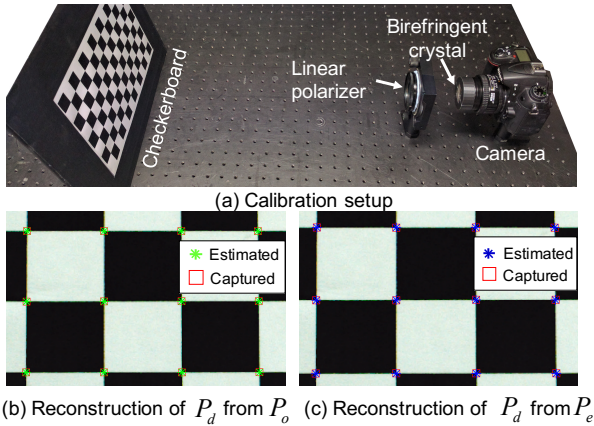
(a) Calibration setup

(b) Reconstruction of P_d from P_o (c) Reconstruction of P_d from P_e

Figure 7: *(a) Our calibration setup for birefringence. We capture three different images of a direct, o-ray and e-ray images. (b) & (c) After finding the optical axis **a**, we obtain the reconstructed positions of P_d from P_o and P_e (see text for details). Our experiment shows an average of just ~0.5 pixel misalignment.*

We therefore devise a novel calibration method, using again the three images of a checkerboard previously obtained to calibrate E. We first extract the position of each corner on the checkerboard for the direct, o-ray, and e-ray images. Let P_d, P_o and P_e be the positions of corresponding pixels for a given corner on the three images, respectively. We can estimate depth z using Equation (1) for the o-ray correspondence between P_d and P_o, since the essential point E has already been calibrated. We now find out the optical axis of the birefringent medium **a**, by parameterizing $\psi_{e \to d}$ with the optical axis **a**, resulting in $\psi_{e \to d}(P_e, z; \mathbf{a})$; this function reconstructs the position of the direct ray from the known values of P_e and z, with an unknown variable **a**. Since we know the ground truth position of the direct ray P_d, we can search for the optimal value of **a** that minimizes the error as:

$$\min_{\mathbf{a}} \sum_{\{P_d, P_e\} \in \Pi} |P_d - \psi_{e \to d}(P_e, z; \mathbf{a})|_2, \qquad (6)$$

where Π is the set of corresponding pairs of P_e and P_d for every corner point. We solve this optimization problem with a constrained nonlinear minimization solver [Waltz et al. 2006]. The average reconstruction error is ~0.5 pixel for each corner as shown in Figure 7.

8 Analysis and Validation

In this section we analyze and evaluate the performance of all the components in our birefractive stereo imaging system.

Calcite Crystal   Calcite is a popular birefringent uniaxial crystal, which shows significant birefringence. It is an ideal material for our imaging system, given its clear transparency. Figure 8(a) shows its transmittance ratio, which increases rapidly from approximately 60% at 400 nm, to a stable 90% across most visible wavelengths. Figure 8(b) presents the indices of refraction for the o-ray and the e-ray, η_o and η_e, respectively; η_o ranges from 1.65 to 1.70 while η_e remains more constant at ~1.48 in the visible spectrum. Figures 8(c) and (d) provide a side-by-side comparison of the color difference between two images, captured with and without the calcite crystal. Despite the lower transmittance of blue light, they show very minor color differences (note that the apparent horizontal blur in (c) is due to the double refraction effect). Figure 8(e) quantitatively compares these color differences using the CIE $u'v'$ chromaticity diagram, showing a very accurate match (the average color
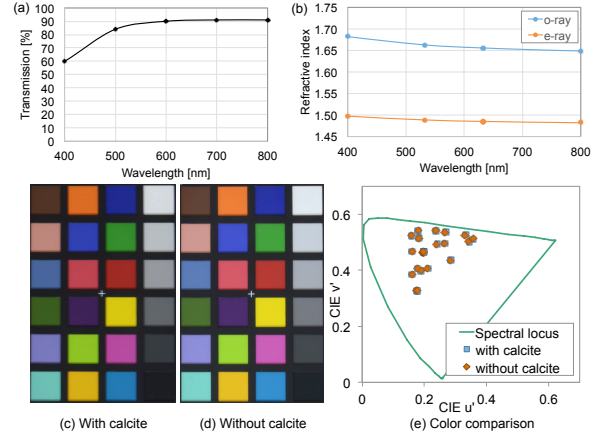

(c) With calcite   (d) Without calcite   (e) Color comparison

Figure 8: *(a) Transmittance ratio across the visible spectrum for calcite, reaching 90% across most visible wavelengths. (b) Refractive indices η_o and η_e for calcite (with a difference of about 0.17 − 0.20). (c) & (d) Photographs of a Color Checker target with and without the crystal, showing good color consistency. (e) Color differences of the two images. The average color difference is only ~0.0016 in the CIE $u'v'$ chromaticity space.*
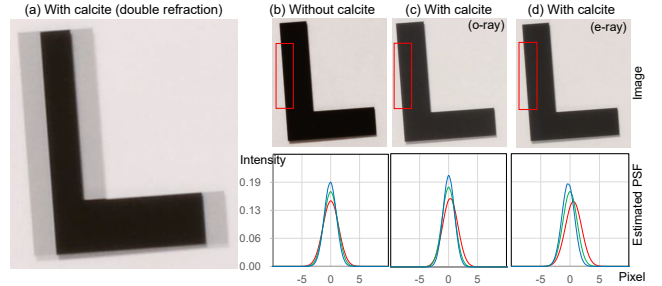

Figure 9: *A Thorlab slant edge target was captured (a) to determine the PSFs of o-ray (c) and e-ray (d) in double refraction [(b) shows the direct image for comparison purposes]. The Gaussian PSFs are estimated from the slanted edges in the red boxes.*

difference is only ~0.0016). We did not experience any pleochroism with the calcite crystal.

Sharpness in Double Refraction   We evaluate the crystal's impact on image sharpness by measuring the point spread functions (PSFs) of the individual o-ray and e-ray. For that purpose, we first capture a Thorlab slant edge target (R2L2S2P) with and without the crystal, as well as through a linear polarizer to isolate the ordinary and the extraordinary refractions, and then estimate the PSFs of the red, green and blue channels for the different images [Burns 2000] (see Figure 9). The per-channel standard deviations are very low for the three images (1.37, 1.32, 1.38 for red, 1.20, 1.13, 1.08 for green, and 1.06, 0.99, 1.08 for blue). While the PSFs without the crystal are reasonably symmetric, the per-channel PSFs of the o-ray and the e-ray images present a shift, due to the wavelength-dependent indices of refraction shown in Figure 8(b). The estimated shifts for the red- and blue-channel PSFs with respect to the green channel are $0.03, -0.06$ for the direct image, $0.28, 0.02$ for the o-ray, and $0.58, -0.31$ for the e-ray. The color shift is smaller than half a pixel, so we can safely ignore the effects of chromatic dispersion in our model.

Crystal Parameters   Here we synthetically analyze the impact of the different parameters of the crystal and the camera on depth resolution. In these experiments, we vary a specific parameter for each plot, while fixing others in the configuration of our prototype. A 45-degree-cut calcite (internal optical axis: 45 degrees) is placed
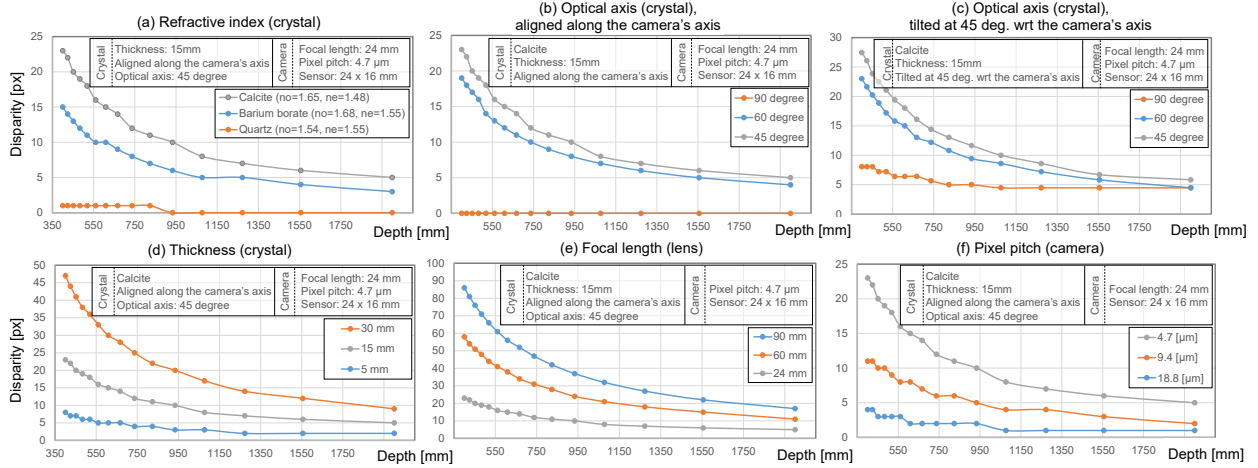
Figure 10: *We evaluate the impact of the optical properties of the crystal on the depth resolution (pixel disparity) in birefractive stereo, including (a) refractive indices, (b) the internal optical axis of the crystal when installed along with the optical axis of the camera and (c) when tilted at an angle of 45 degrees with respect to the camera, and (d) crystal thickness. In addition, we also evaluate (e) the impact of the focal length of the lens, and (f) the pixel pitch of the camera.*
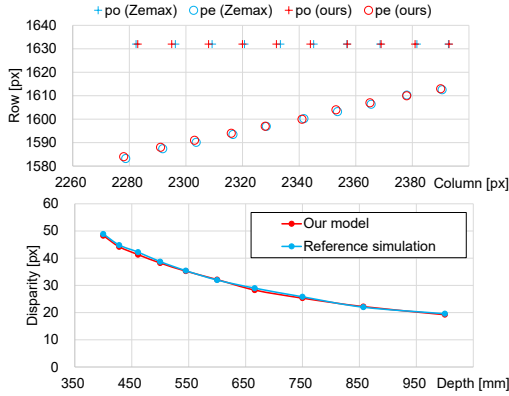

Figure 11: *Validation of our image formation model using the Zemax simulation software, for a depth range of 40 – 100 cm (optical axis $\mathbf{a} = [-0.18, 0.43, 0.88]^T$, and tilting the calcite crystal at an angle of 45 degrees along the y-axis). Our image formation shows a strong agreement in both the positions (top) and the disparity (bottom) of the two refracted rays.*

along the optical axis of the camera (pixel pitch: 4.7 μm) equipped with a 24 mm lens. First, we analyze the choice of calcite as the birefringent medium. There are other birefringent crystals commercially available with a high transmission, which could be used in principle. Figure 10(a) shows disparity using calcite, barium borate, and quartz. Since larger birefringence increases depth resolution, calcite becomes the optimal choice given its larger difference between its two indices of refraction (0.17 for calcite, 0.13 for barium borate, 0.01 for quartz). We compare three optical axes: 45, 60 and 90 degrees by placing the crystal aligned with the camera [Figure 10(b)], and tilted at an angle of 45 degrees [Figure 10(c)]. Regardless of the angle between the crystal and the camera, a 45-degree optical axis provides the highest depth resolution. Note that a crystal with a 90-degree optical axis loses its depth discrimination power when aligned with the camera. Figure 10(d) shows the effect of thickness. For the three thicknesses tested, the thickest medium achieves higher depth resolution by displacing the o- and e-rays more.

Imaging Parameters In addition to the crystal parameters, we analyze the effect of the focal length and pixel pitch on depth resolution. Note that the distance between the birefringent medium and

the center of projection does not affect depth resolution, since the incoming and outgoing rays at the medium have the same direction due to the parallelism between the surfaces of the medium, as shown in Figure 3(a). Focal length is linearly correlated with depth resolution [Equation (3)] as shown in Figure 10(e). As for the sensor resolution, since we rely on pixel displacement to distinguish depth, more pixels lead to increased depth resolution, as shown in Figure 10(f).

Model Validation Last, we synthetically validate our proposed image formation model using the Zemax ray simulation software, popularly used in optics. Given a pixel on the image plane, our model yields pixel positions P_o and P_e for the o- and e-rays on the sensor plane. We explore a set of depth candidates ranging from 40 cm to 100 cm, and compare the resulting disparities with the results obtained from Zemax. This comparison is plotted in Figure 11, showing a strong agreement between our model and the professional optics software.

9 Results

We employed two different calcite crystals of the same size (25×25×15 mm), but with different optical axes of 45 and 90 degrees respectively. For the 45-degree-cut calcite, we place the crystal right in front of the lens compactly, so that the crystal's normal is aligned with the camera's optical axis, as shown in Figure 1(a). For the 90-degree-cut calcite, the crystal is slanted at an angle of 45 degrees with respect to the camera's optical axis to retain enough disparity, as shown in Figure 10(c). We use a Nikon D7000 camera to capture linear RGB images at a resolution of 4928×3264 (pixel pitch: 4.7 μm). Either a Nikon 24 mm lens or a CoastalOpt apochromatic lens of 60 mm are used, depending on the depth range of the scene, with a fixed aperture size of $f/11$ in general. Our Matlab and C++ implementation runs on an Intel i7 3.40 GHz with 32 GB RAM without GPU acceleration. Geometric calibration takes approximately two seconds, excluding capture times, while building the birefractive disparity tables takes about a minute. As for depth estimation, it takes a little over two minutes, including all the steps: edge extraction (10 s), computation of the matching costs (55 s), cost aggregation (50 s), and depth propagation (25 s). In the following we show our final results, compared to other state-of-the-art techniques, as well as applications for refocusing, anaglyph generation, background-foreground segmentation and recoloring, and depth-aware image compositing.
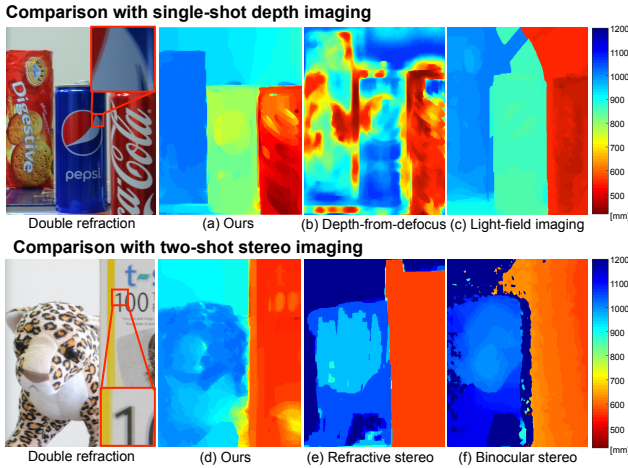
Comparison with single-shot depth imaging

Double refraction   (a) Ours   (b) Depth-from-defocus   (c) Light-field imaging

Comparison with two-shot stereo imaging

Double refraction   (d) Ours   (e) Refractive stereo   (f) Binocular stereo

Figure 12: *Top: Comparison with other single-shot depth imaging methods: (a) Our method, (b) Chakrabarti and Zickler [2012], and (c) Lytro light-field camera. Bottom: Our method also produces good results compared with two-shot stereo methods: (d) Our method, (e) Chen et at. refractive stereo [2013], and (f) Hirschmuller binocular stereo [2005].*

(a) Double refraction   (b) Ours   (c) Reflective stereo

Figure 13: *Comparison of our birefringence image formation model with the double-reflection model of Yano et al. [2010]. (a) Input double-refraction image. (b) Result using our model. (c) Result using the double-reflection model.*

Comparison with Other Approaches We first compare our method with other single-shot approaches: a depth-from-defocus (DfD) method [Chakrabarti and Zickler 2012], and a light-field camera (Lytro Illum) [Figures 12(a) – (c)]. The DfD method suffers from low accuracy due to edge smoothness ambiguity (b). The light-field camera (c) produces better results than DfD, although the spatial resolution needs to be lowered significantly to accommodate the increased angular resolution.

In Figures 12(d) – (f), we additionally compare our results (d) with other traditional methods that require at least two shots: refractive stereo [Chen et al. 2013] (e), and binocular stereo [Hirschmuller 2005] (f). Despite relying on a single-shot, the depth quality of our method is highly competitive compared to two-shot stereo methods.

Compared to reflective stereo [Yano et al. 2010], our technique does not require a large form factor for the imaging setup. While Yano's work approximates the linear projection of continuously attenuated images using a Taylor expansion, our method calculates dual-matching cost volumes in the gradient domain, accounting for the physical nature of birefringence (Section 6). We nevertheless test if the imaging model in reflective stereo could be used with our double-refraction image; as Figure 13 demonstrates, our birefringence image formation model produces superior results.

Depth of Field The performance of birefractive stereo is limited by defocus blur, which can degrade the double refraction effect. Figure 14 compares depth maps of a scene by varying the
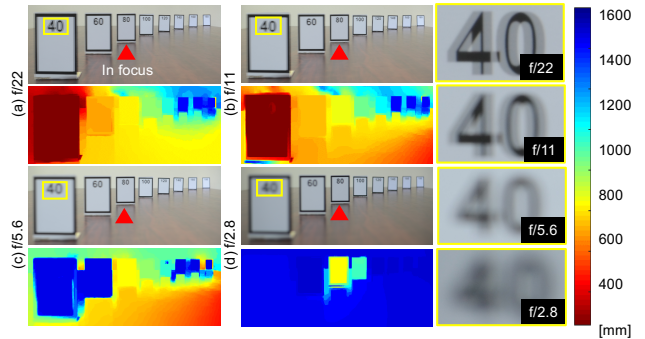

Figure 14: *The impact of the aperture size on depth estimation. (a) & (b) Captures with small apertures, $f/22$ and $f/11$, can estimate depth in the range of 40 cm to 160 cm successfully, (c) & (d) while captures with larger apertures $f/5.6$ and $f/2.8$ fail as double refraction is damaged by blur.*
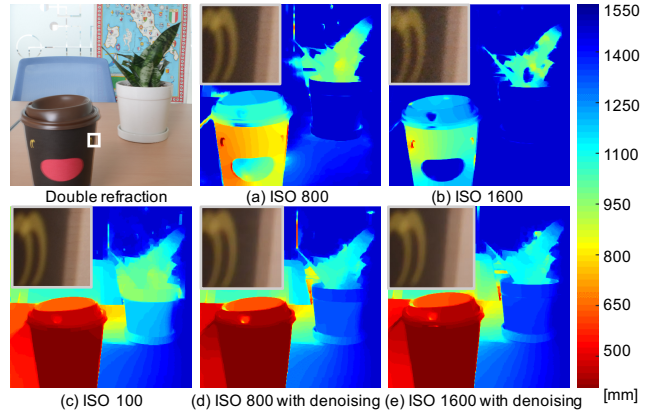

Figure 15: *The impact of noise on depth estimation: (a) & (b) Severe noise in high ISOs leads to suboptimal depth estimation; on the other hand, (c) a low ISO yields better performance due to the reliable matching between pixel gradients. (d) & (e) To mitigate the effect of noise, we can apply a simple denoising algorithm, which leads to significantly improved results.*

f-number. Small apertures, such as $f/22$ and $f/11$, retain this double refraction, allowing us to recover depth in the range of 40 cm to 160 cm. However, larger apertures of $f/5.6$ and $f/2.8$ damage double refraction by blur, due to the shallow depth of field. To overcome this limitation, we could combine depth-from-defocus techniques with our method.

Image Noise We investigate the impact of noise on depth estimation. Figure 15 compares depth maps estimated by varying the ISO, i.e., the shutter speed is adjusted to have the same exposure levels when changing the ISO. While lower ISOs allow reliable depth estimates, higher ISOs lead to suboptimal results. We can mitigate this problem by simply applying a 5×5 median filter to the noisy inputs before computing correspondences.

10 Applications

Refocusing Figure 17 shows a refocusing application. We first estimate a depth map from a double-refraction image using the model presented in this paper. Our image formation allows us to build a set of deconvolution kernels for each depth; an example is shown in the inset, consisting of the two point spread functions for the o-ray and the e-ray. By applying different blur kernels across the estimated depths, we can refocus the captured photograph, as Figures 17(d) and (e) show.
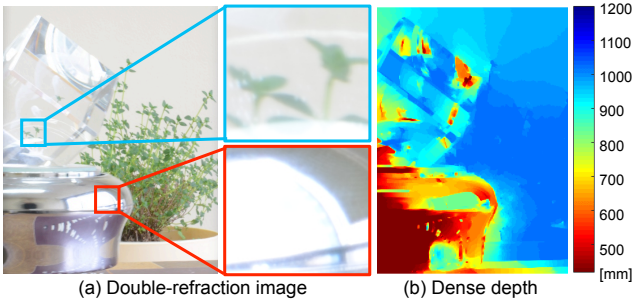
Figure 16: *Failure case: Inheriting the typical limitations of existing stereo algorithms, our method does not handle transparent or highly reflective objects well, leading to wrong depth estimations.*
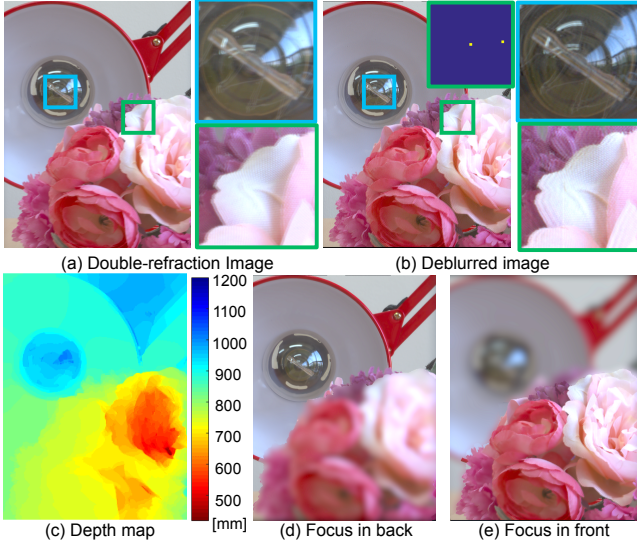

Figure 17: *Refocusing application. From (a) the double-refraction image, we restore (b) a sharp image using non-blind Lucy-Richardson deconvolution with (c) the estimated depth map. An example kernel is shown in the inset of (b). Using the kernels and the depth map, we can perform synthetic refocusing, focusing on (d) the red light, and on (e) the flowers.*

Anaglyph Generation We generate stereo pairs from a single image by using our depth estimation to render novel views, from which we create an anaglyph (see Figure 18). The bottom row shows the newly generated images, as well as close-up details.

Object Segmentation Using our estimated depth and the underlying color information, we can quickly perform object segmentation in images using existing techniques like grab-cut [Rother et al. 2004]. Figure 19 shows segmentation examples. Subsequent effects like the decolorization shown become straightforward.

Depth-Aware Image Compositing Our technique enables automatic depth-aware image compositing, where occlusions are directly handled by our estimated depth without user intervention. In Figure 20, we take the cat extracted in the previous example, and automatically place it between the can and the teddy bear.

11 Discussion

We have presented a novel birefractive stereo method, based on our image formation model for double refraction, which enables single-shot depth acquisition in a cost-effective and compact form. In addition, we have provided a novel calibration method for the in-
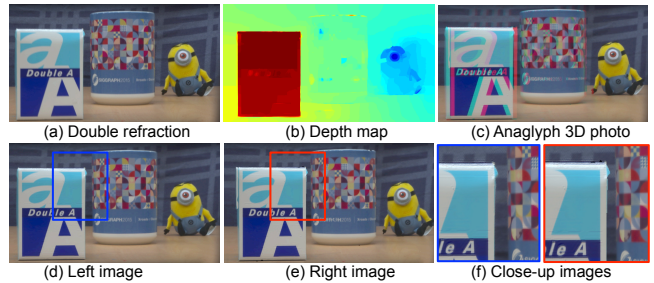

Figure 18: *3D anaglyph image creation using our birefractive stereo. Refer to the supplemental material for enlarged figures.*
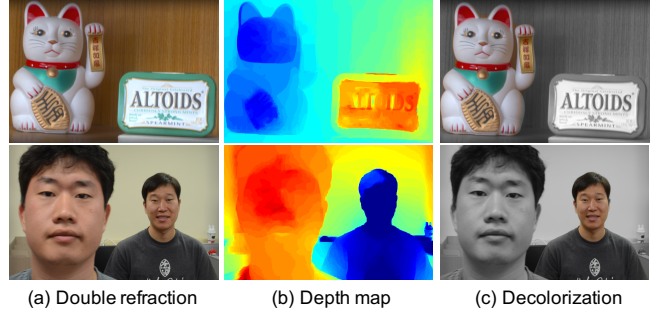

Figure 19: *Automatic object segmentation and depth-aware decolorization using our technique.*
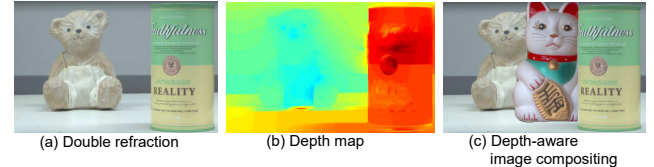

Figure 20: *Occlusion-aware image compositing using our depth map from birefractive stereo.*

trinsic and extrinsic properties of birefringent crystals, and demonstrated several useful image-editing applications that leverage our estimated depth. Since calcite is a very transparent medium, placing it in front of the lens does not degrade visibly the captured images; neither light loss nor color variations become problematic; nevertheless, we argue that this is an acceptable trade-off, given the depth information we obtain and the ease of use of our approach.

We note that an image captured through a coded aperture with two holes would be similar to our double-refraction image using a birefractive medium. However, such coded aperture decreases light throughput significantly, while the maximum disparity is limited by the diameter of the aperture.

Limitations & Future Work There are still some interesting avenues for future work to improve our system: As most stereo algorithms for depth acquisition, transparency and specular reflections are difficult to handle leading to wrong depth estimations (see Figure 16). Our image formation model assumes that the refractive medium consists of two planar parallel faces. Commercially available birefringent crystals comply sufficiently well with this requirement, as our results show; however, our method is not applicable to arbitrary shapes. In addition, high-frequency repeated patterns on a scene could be misinterpreted as double refraction effects resulting in failure of depth estimation. Even though tabulation for mapping correspondences is a sensible alternative, developing efficient analytic formulas for the mapping functions remains as future work. Combining depth-from-defocus with birefractive stereo could help overcome the limitations due to the presence of blur. Also, despite the good results obtained with denoising algorithms, noise-robust correspondence matching techniques could be investigated. Last, we assume that the intensities of the two refracted rays are the same,

although differently polarized rays often show different levels of intensity due to the Fresnel effect at grazing angles. Identifying such intensity changes from our single image with overlapped refraction information remains as an open, challenging problem.

Acknowledgements

References

AVENDAÑO-ALEJO, M., STAVROUDIS, O. N., AND Y GOITIA, A. R. B. 2002. Huygenss principle and rays in uniaxial anisotropic media. i. crystal axis normal to refracting surface. *JOSA A 19*, 8, 1668–1673.

BAEK, S.-H., AND KIM, M. H. 2015. Stereo fusion using a refractive medium on a binocular base. In *Proc. Asian Conference on Computer Vision (ACCV 2014)*, Springer, Singapore, Singapore, vol. 9004 of *LNCS*, 503–518.

BAEK, S.-H., AND KIM, M. H. 2016. Stereo fusion: Combining refractive and binocular disparity. *Computer Vision and Image Understanding 146*, 52–66.

BANDO, Y., CHEN, B.-Y., AND NISHITA, T. 2008. Extracting depth and matte using a color-filtered aperture. In *ACM Transactions on Graphics (TOG)*, vol. 27, ACM, 134.

BARRON, J. T., AND MALIK, J. 2013. Intrinsic scene properties from a single rgb-d image. In *Proc. Comput. Vision and Pattern Recognition (CVPR)*, 17–24.

BURNS, P. D. 2000. Slanted-edge MTF for digital camera and scanner analysis. In *Proc. the Conference on Image Processing, Image Quality, Image Capture Systems (PICS-00)*, 135–138.

CHAKRABARTI, A., AND ZICKLER, T. 2012. Depth and deblurring from a spectrally-varying depth-of-field. In *European Conference on Computer Vision (ECCV), 2012*, Springer, 648–661.

CHEN, Z., WONG, K.-Y. K., MATSUSHITA, Y., AND ZHU, X. 2013. Depth from refraction using a transparent medium with unknown pose and refractive index. *Int. J. Comput. Vision (ICJV)*, 1–15.

GAO, C., AND AHUJA, N. 2006. A refractive camera for acquiring stereo and super-resolution images. In *Proc. Comput. Vision and Pattern Recognition (CVPR)*, 2316–2323.

GUY, S., AND SOLER, C. 2004. Graphics gems revisited: Fast and physically-based rendering of gemstones. *ACM Trans. Graph. 23*, 3 (Aug.), 231–238.

HECHT, E. 2002. Optics, 4th. *International edition, Addison-Wesley, San Francisco 3*.

HIRSCHMULLER, H. 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. Comput. Vision and Pattern Recognition (CVPR)*, 807–814.

LATORRE, P., SERON, F. J., AND GUTIERREZ, D. 2012. Birefringence: calculation of refracted ray paths in biaxial crystals. *The Visual Computer 28*, 4, 341–356.

LEE, D., AND KWEON, I. 2000. A novel stereo camera system by a biprism. *IEEE Trans. Robotics and Automation 16*, 5, 528–541.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. *ACM Trans. Graph. (TOG) 23*, 3, 689–694.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. In *ACM Transactions on Graphics (TOG)*, vol. 26, ACM, 70.

LI, X., AND WANG, R. 2009. Analysis and optimization of the stereo system with a biprism adapter. In *International Conference on Optical Instrumentation and Technology*, International Society for Optics and Photonics, 75061V1–75061V8.

LIANG, Q.-T. 1990. Simple ray tracing formulas for uniaxial optical crystals. *Applied Optics 29*, 7, 1008–1010.

ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. 23*, 3 (Aug.), 309–314.

SHESTAK, S., KIM, D.-S., AND CHA, K.-H. 2015. Stereoscopic three-dimensional television using active glasses with switchable refraction. *Journal of Electronic Imaging 24*, 3, 033006–033006.

SHI, J., TAO, X., XU, L., AND JIA, J. 2015. Break ames room illusion: Depth from general single images. *ACM Trans. Graph. 34*, 6 (Oct.), 225:1–225:11.

SHIMIZU, M., AND OKUTOMI, M. 2006. Reflection stereo-novel monocular stereo using a transparent plate. In *Proc. Canadian Conf. Computer and Robot Vision (CRV)*, IEEE, 14–14.

TANNENBAUM, D., TANNENBAUM, P., AND WOZNY, M. 1994. Polarization and birefringency considerations in rendering. In *Computer Graphics. Proceedings, Annual Conference Series, SIGGRAPH*, ACM, 221–222.

TSAI, T.-H., AND BRADY, D. J. 2013. Coded aperture snapshot spectral polarization imaging. *Applied optics 52*, 10, 2153–2161.

WALTZ, R. A., MORALES, J. L., NOCEDAL, J., AND ORBAN, D. 2006. An interior algorithm for nonlinear optimization that combines line search and trust region steps. *Mathematical programming 107*, 3, 391–408.

WEIDLICH, A., AND WILKIE, A. 2008. Realistic rendering of birefringency in uniaxial crystals. *ACM Transactions on Graphics (TOG) 27*, 1, 6.

YANG, Q. 2012. A non-local cost aggregation method for stereo matching. In *Proc. Comput. Vision and Pattern Recognition (CVPR)*, 1402–1409.

YANO, T., SHIMIZU, M., AND OKUTOMI, M. 2010. Image restoration and disparity estimation from an uncalibrated multi-layered image. In *Proc. Comput. Vision and Pattern Recognition (CVPR)*, IEEE, 247–254.

ZALEVSKY, Z., AND BEN-YAISH, S. 2007. Extended depth of focus imaging with birefringent plate. *Optics express 15*, 12, 7202–7210.

ZHANG, L., AND NAYAR, S. K. 2006. Projection Defocus Analysis for Scene Capture and Image Display. *ACM Trans. on Graphics* (Jul).

ZHANG, Z. 2000. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 22*, 11, 1330–1334.