HAL
open science

Multiple-View Geometry of the Refractive Plane

Visesh Chari, Peter Sturm

To cite this version:

Visesh Chari, Peter Sturm. Multiple-View Geometry of the Refractive Plane. BMVC 2009 - 20th British Machine Vision Conference, Sep 2009, London, United Kingdom. pp.1-11, 10.5244/C.23.56 . inria-00434342

HAL Id: inria-00434342

https://inria.hal.science/inria-00434342v1

Submitted on 24 May 2011

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

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

Multi-View Geometry of the Refractive Plane

Visesh Chari and Peter Sturm

Abstract

Transparent refractive objects are one of the main problems in geometric vision that have been largely unexplored. The imaging and multi-view geometry of scenes with transparent or translucent objects with refractive properties is relatively less well understood than for opaque objects. The main objective of our work is to analyze the underlying multi-view relationships between cameras, when the scene being viewed contains a single refractive planar surface separating two different media. Such a situation might occur in scenarios like underwater photography. Our main result is to show the existence of geometric entities like the fundamental matrix, and the homography matrix in such instances. In addition, under special circumstances we also show how to compute the relative pose between two cameras immersed in one of the two media.

1 Introduction

The multi-view geometry resulting from opaque scenes is now well understood, for the case of perspective projection. To some extent, even the insertion of reflective elements has been studied in the area of catadioptrics [14, 13]. The phenomenon of refraction, however, has largely been left un-addressed in the vision community.

The introduction of refractive elements into a scene changes the multi-view geometry that results from the imaging phenomenon. Until recently, the study of this change had been restricted to multimedia photogrammetry [7] and oceanic engineering [9], where the major perspectives were to either neglect the refraction [3, 9], view refraction as an error or aberration to perspective imaging [12, 10], or to look at its correction as an iterative optimization problem [7]. In computer vision, some of the first attempts have come in the recent past [5, 8, 1, 15, 6]. Kutulakos et al. [5] investigate the geometry of *light-path triangulation*, which aims to find conditions and algorithms where reconstruction of individual tracks of light is possible. The work of Morris and Kutulakos [8] looks at refractive stereo. The main idea here is to estimate the normal of the refractive surface, given 2D-3D correspondences, irrespective of the refractive index. They argue that under the assumption that light is refracted only once, two views are sufficient to reconstruct an arbitrary refractive surface. The work of Nayar et al. [1] concentrates on the estimation of the shape of the object by optimizing a function that minimizes the difference between observed and predicted images, based on a suitable parametric model of the object shape, as well as assuming an affine setting. The works of Singh et al. [15] and Thierry et al. [6] concentrate on the internal calibration of cameras underwater, when a planar refractive surface separates the two media (air/water).

However, to the best of our knowledge, the epipolar geometry resulting from refraction has not been analyzed till now. In the presence of refraction, this geometry not only encodes the relative position of the cameras, but also information about the relative refractive index between the two

media in question, and the geometry of the surface separating the concerned media. In this paper, we concentrate on the multi-view relationships induced by a refractive planar surface. We choose to deviate from the generic scene assumption of Kutulakos *et al.* [5], which enables us to completely describe the epipolar geometry in the presence of refraction. In this respect, our work is closer to Barreto & Sturm [13] and Glaeser & Schröcker [2]. In [2], the authors define the equation that governs the projection of a 3D point onto an image, while being refracted across a planar surface. We follow an approach that is along the lines of [13] in deriving the various multi-view relationships in the presence of refractions.

In the following sections, we derive the various multi-view relationships between two views of a scene, when a planar transparent surface separates the scene from the camera. The two media are assumed to have different refractive indices. First, we define the projection matrix for such cameras for lines, where we show that a 3D line is mapped to a quartic curve. Next we derive the *fundamental* matrix between two views, which turns out to be a function of the camera poses and the plane normal and refractive index. We then derive an expression for the homography between a scene plane and its image on the camera. Finally, we give algorithms for obtaining the relative pose between two cameras in specialized cases, by deriving the expression of what is commonly referred to as *Snell's Window*.

Background & Notation Consider the scenario where the scene consists of a planar transparent refractive surface, with cameras on one side of the plane. Let the relative refractive index between the two media on either side of the plane be denoted by λ ($= \frac{n_1}{n_2}$, where n_1 is in the medium with cameras). Let us remember that the incident and refracted angles are related by

$$\sin(\theta_2) = \lambda \sin(\theta_1) \tag{1}$$

where θ_1 is the incident angle, and θ_2 is the refracted angle.

Images of objects on the opposite side of the cameras are formed after refraction by the surface in consideration. We wish to deduce external calibration information about the cameras and the geometry of the refracting plane from these set of images alone. Further along this document, we make certain assumptions about the geometry of the scene, refractive index etc. to simplify matters. Such assumptions will be detailed when necessary.

In the rest of the document, we use the following coordinate system. Let the 3D refractive plane, denoted by $\pi = \begin{pmatrix}\mathbf{v}^\top & d\end{pmatrix}^\top$ take the values $\begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}^\top$, which aligns it with the X-Y plane of the world coordinate system. Let two cameras P_1 and P_2 be situated in the following manner

$$\mathsf{P}_1 = \mathsf{R}_1 \begin{pmatrix}\mathsf{I} & -\mathbf{t}_1\end{pmatrix} \tag{2}$$

$$\mathsf{P}_2 = \mathsf{R}_2 \begin{pmatrix}\mathsf{I} & -\mathbf{t}_2\end{pmatrix} \tag{3}$$

$$\mathbf{t}_1 = \begin{pmatrix}0 & 0 & -1\end{pmatrix}^\top \tag{4}$$

where $\mathsf{R}_1, \mathsf{R}_2$ are rotation matrices (R if there is only one camera) and I is the identity matrix.

Now, the only undecided component about the world coordinate system is the location of the X- and Y-axes, which can vary upto rotation about the Z-axis. Although this can easily be fixed by assuming the first camera's coordinate axes to be aligned with that of the world, we ignore this unnecessary assumption for now. Finally, for the purposes of this document, the internal parameters of all cameras are assumed to be $\mathsf{I}_{3\times3}$,i.e., the cameras are calibrated.

Now, let us define notations for points. We denote an image point by the variable $\mathbf{x}$, which is a 3×1 homogeneous vector defined as $\mathbf{x} = \begin{pmatrix}x & y & 1\end{pmatrix}^\top$. The direction vector for an image

point is a 4×1 homogeneous coordinate and is represented by $\mathbf{q} = \begin{pmatrix}\tilde{\mathbf{q}}_{3\times1} & 0\end{pmatrix}^\top = \begin{pmatrix}(\mathsf{R}^\top\tilde{\mathbf{x}})^\top & 0\end{pmatrix}^\top$, and its jth coordinate for the ith camera is represented by $q_{i,j}$, or by q_j when there is only one camera. The vector $\tilde{\mathbf{x}}$ represents the unit vector corresponding to image point $\mathbf{x}$. Thus $\|\tilde{\mathbf{q}}\| = 1$. A vector formed by collecting elements from another is represented in braces. For example, $\mathbf{L}_{1,(6,1,2)}$ represents a vector formed by elements $6, 1$ and 2 in $\mathbf{L}_1$. The Plücker coordinates of a 3D line is denoted by $\mathbf{L}$. Matrices are represented using the font M, and vectors are represented in bold ($\mathbf{v}$). ($\widehat{\mathbf{q}} = \begin{pmatrix}q_1^2 & q_1q_2 & q_2^2 & q_1q_3 & q_2q_3 & q_3^2\end{pmatrix}^\top$ [13, 4]) denotes *lifted* coordinates of any vector $\mathbf{q}$. Finally, if two vectors are related by a linear transformation ($\mathbf{q} = \mathsf{T}\tilde{\mathbf{x}} = \begin{pmatrix}\mathsf{R}^\top & 0_{3\times1}\end{pmatrix}^\top \tilde{\mathbf{x}}$), their lifted coordinates are related by [13, 4]

$$\widehat{\mathbf{q}} = \mathsf{D}^{-1}\mathsf{S}(\mathsf{T} \otimes \mathsf{T})\mathsf{S}^\top\widehat{\tilde{\mathbf{x}}} \tag{5}$$

where $\otimes$ is the Kronecker product, and D, S are matrices used to re-arrange rows and for numerical division by constants. Henceforth, all derivations are done with $\mathbf{q}$, while noting that the corresponding equations for $\mathbf{x}$ can be obtained by using Equation 5.

2 The Camera Projection Matrix

The first step in deriving the epipolar geometry needed to describe multi-view relationships is the external camera matrix. We derive the camera matrix by first defining the back-projection of a point, and then focusing on the condition that it intersects with a 3D line.

2.1 Back-projecting a Point

It has been previously shown that the image of a 3D point after refraction from a plane is the solution of a quartic equation that lies within a specified interval [2]. We proceed to show that the image of a line after refraction is a quartic curve, and from this projection we derive a description of the camera matrix for refraction.

Consider a 3D point with inhomogeneous coordinates $\mathbf{X}$. An image point $\mathbf{x}$ with 3D direction ray $\mathbf{q}$ (both in homogeneous coordinates) represents an image corresponding to $\mathbf{X}$, if the line from the camera centre in the direction of $\mathbf{q}$ passes through $\mathbf{X}$ after refraction. After developing the refraction equation 1 for vectors, we can get the point $\mathbf{Q}$ on the refraction plane π, and the direction of the refracted ray $\mathbf{R}$ as

$$\mathbf{Q} = \begin{pmatrix} t_3\tilde{\mathbf{q}} - q_3\mathbf{t} \\ -q_3 \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} q_1\lambda \\ q_2\lambda \\ \sqrt{(1-\lambda^2) + \lambda^2 q_3^2} \\ 0 \end{pmatrix} \tag{6}$$

where $Q_3 = 0$. The corresponding Plücker coordinates of the line can be written as

$$\mathbf{L} \;=\; \begin{pmatrix} Q_4\mathbf{R}_{1\ldots3} - R_4\mathbf{Q}_{1\ldots3} \\ \mathbf{Q}_{1\ldots3} \times \mathbf{R}_{1\ldots3} \end{pmatrix} = \begin{pmatrix} -q_1q_3\lambda \\ -q_2q_3\lambda \\ -q_3\sqrt{(1-\lambda^2) + \lambda^2 q_3^2} \\ (q_2t_3 - q_3t_2)\sqrt{(1-\lambda^2) + \lambda^2 q_3^2} \\ (q_3t_1 - t_3q_1)\sqrt{(1-\lambda^2) + \lambda^2 q_3^2} \\ (q_1t_2 - q_2t_1)q_3\lambda \end{pmatrix} \tag{7}$$

Now, if $\mathbf{L}$ intersects a 3D line, it can be concluded that this 3D line projects onto the points given by the locus of $\mathbf{q}$, and hence $\mathbf{x}$. Using this principle, we derive the projection matrix.

2.2 The Refractive Projection Matrix

Let $\mathbf{A}, \mathbf{B}$ be two homogeneous points on the 3D line. The Plücker coordinates of this line are given as $\mathbf{L}_1 = \begin{pmatrix} A_4\mathbf{B}_{(1,2,3)} - B_4\mathbf{A}_{(1,2,3)} & \mathbf{A}_{(1,2,3)} \times \mathbf{B}_{(1,2,3)} \end{pmatrix}^\top$. The line $\mathbf{L}$ in Equation 7 intersects this line if and only if these two lines satisfy the Klein quadric [11], defined by the matrix $\mathsf{W} = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}$. This can be developed as follows

$$\mathbf{L}_1^\top \mathsf{W} \mathbf{L} = 0 \tag{8}$$

$$\begin{pmatrix} \mathbf{L}_{1,(4,5,6)}^\top & \mathbf{L}_{1,(1,2,3)}^\top \end{pmatrix} \mathbf{L} = 0 \tag{9}$$

$$\left(\mathbf{L}_{1,(6,1,2)}\right)^\top \begin{pmatrix} -q_3 \\ q_2t_3 - q_3t_2 \\ q_3t_1 - t_3q_1 \end{pmatrix} \sqrt{(1-\lambda^2) + \lambda^2 q_3^2} = q_3\lambda \left(\mathbf{L}_{1,(4,5,3)}\right)^\top \begin{pmatrix} q_1 \\ q_2 \\ q_2t_1 - q_1t_2 \end{pmatrix} \tag{10}$$

Squaring both sides removes the square root. Noting that for two vectors $\mathbf{a}$ and $\mathbf{b}$, $(\mathbf{a}\cdot\mathbf{b})^2 = \mathsf{D}\widehat{\mathbf{a}} \cdot \widehat{\mathbf{b}}$ or $\widehat{\mathbf{a}} \cdot \mathsf{D}\widehat{\mathbf{b}}$ where $\mathsf{D} = \operatorname{diag}\begin{pmatrix} 1 & 2 & 1 & 2 & 2 & 1 \end{pmatrix}$, we get

$$\left(\widehat{\mathbf{S}}(1-\lambda^2+\lambda^2q_3^2) - \widehat{\mathbf{T}}\lambda^2q_3^2\right)\mathsf{D_s}\widehat{\mathbf{q}} = 0 \tag{11}$$

$$\begin{pmatrix} \widehat{\mathbf{S}}(1-\lambda^2) & \lambda^2(\widehat{\mathbf{S}} - \widehat{\mathbf{T}}) \end{pmatrix} \underbrace{\begin{pmatrix} \mathsf{D_s} & 0 \\ 0 & \mathsf{D_s} \end{pmatrix}}_{\bar{\mathsf{D}}_s} \begin{pmatrix} \widehat{\mathbf{q}} \\ q_3^2\widehat{\mathbf{q}} \end{pmatrix} = 0 \tag{12}$$

$$\mathbf{S} = \left(\mathbf{L}_{1,(6,1,2)}\right)^\top \underbrace{\begin{pmatrix} 0 & 0 & -1 \\ 0 & t_3 & -t_2 \\ -t_3 & 0 & t_1 \end{pmatrix}}_{\mathsf{t_s}}, \mathbf{T} = \left(\mathbf{L}_{1,(4,5,3)}\right)^\top \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t_2 & t_1 & 0 \end{pmatrix}}_{\mathsf{t_t}} \tag{13}$$

$\widehat{\mathbf{S}}$ and $\widehat{\mathbf{T}}$ are defined as functions of $\mathsf{t_s}$ and $\mathsf{t_t}$ respectively, using Kronecker products and symmetric vectorization as defined in Equation 5 (See also [13, 4]).

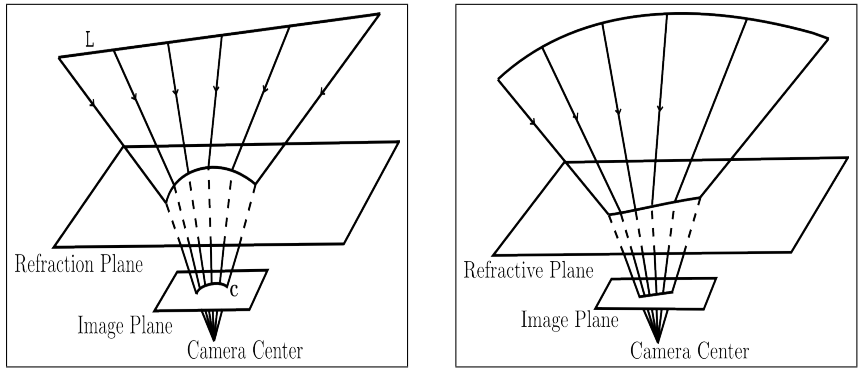
Observe that equation 12. is a quartic curve in the image coordinates (since the lifted coordinates of $\mathbf{q}$ and $\mathbf{x}$ in Equation 5 are related by linear transformations). We can thus conclude that a 3D line is imaged to a quartic curve in a perspective camera, if seen through a refractive planar surface. Notice that the coefficients of the quartic curve in $\mathbf{q}$ can be defined as a linear function of $\widehat{\mathbf{L}_1}$. Thus, if we have a vector $\mathbf{c}$ such that $\mathbf{c}^\top \begin{pmatrix} \widehat{\mathbf{q}} \\ \widehat{\mathbf{q}}q_3^2 \end{pmatrix} = 0$, then $\mathbf{c}$ can be written as

$$\mathbf{c} = \bar{\mathsf{D}}_s^\top \begin{pmatrix} \widehat{\mathbf{S}}^\top(1-\lambda^2) \\ \lambda^2(\widehat{\mathbf{S}} - \widehat{\mathbf{T}})^\top \end{pmatrix} \tag{14}$$

$$= \underbrace{\bar{\mathsf{D}}_s^\top \begin{pmatrix} (1-\lambda^2)\mathsf{D_s}^{-1}\mathsf{S_s}\mathsf{t}_s \otimes \mathsf{t}_s\mathsf{S_s}^\top & 0 \\ \lambda^2\mathsf{D_s}^{-1}\mathsf{S_s}\mathsf{t}_s \otimes \mathsf{t}_s\mathsf{S_s}^\top & -\lambda^2\mathsf{D_t}^{-1}\mathsf{S_t}\mathsf{t}_t \otimes \mathsf{t}_t\mathsf{S}_t^\top \end{pmatrix}}_{\mathsf{P}} \begin{pmatrix} \widehat{\mathbf{L}}_{1,(6,1,2)} \\ \widehat{\mathbf{L}}_{1,(4,5,3)} \end{pmatrix} \tag{15}$$

where P is the projection matrix (Equation 15).

We have thus derived the projection matrix for a perspective camera viewing a scene separated by a planar refracted surface. It projects the lifted Plücker coordinates of a line, $\mathbf{L}_1$, onto a quartic curve in the image space, whose coefficients are elements of the vector $\mathbf{c}$ (Figure 1a).



(a) (b)

Figure 1: (a) shows the projection of a 3D line onto the camera after refraction. (b) shows the back-projection of a line.

3 The Refractive Fundamental Matrix

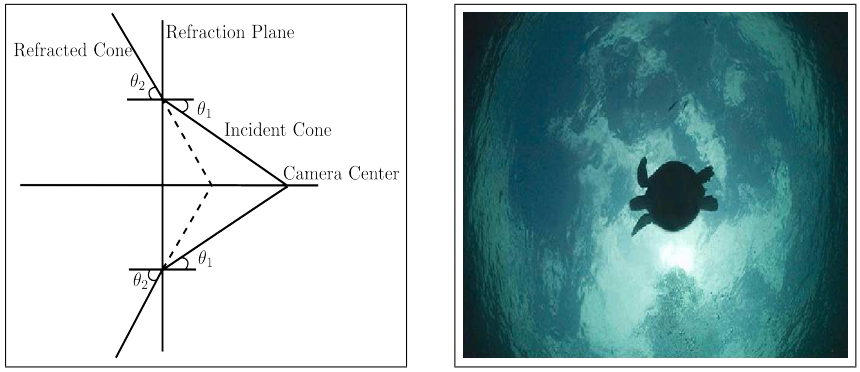
In the previous section, we defined the projection matrix for 3D lines. Continuing the same argument, we can see that the epipolar curve in the first image is nothing but the projection of a line. This line, in turn, is the back-projection of an image point from the second camera. Using the back-projection equation 7, we get the back-projected ray from the second image as

$$\mathbf{L}_2 = \begin{pmatrix} -q_{2,1}q_{2,3}\lambda \\ -q_{2,2}q_{2,3}\lambda \\ -q_{2,3}\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,2}t_{2,3}-q_{2,3}t_{2,2})\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,3}t_{2,1}+t_{2,3}q_{2,1})\sqrt{1-\lambda^2(1-q_{2,3}^2)} \\ (q_{2,1}q_{2,3}t_{2,2}-q_{2,2}q_{2,3}t_{2,1})\lambda \end{pmatrix} \tag{16}$$

where $\mathbf{q}_2 = \begin{pmatrix} q_{2,1} & q_{2,2} & q_{2,3} \end{pmatrix}^\top$ and $\mathbf{t}_2 = \begin{pmatrix} t_{2,1} & t_{2,2} & t_{2,3} \end{pmatrix}$ are the image point and the translation of the second camera. Substituting $\mathbf{L}_2$ in the place of $\mathbf{L}_1$ in Equation 15 will give us the desired result. To do so, we first observe that

$$\widehat{\mathbf{L}_{2,(6,1,2)}} = \mathsf{D}_u^{-1}\mathsf{S}_u\mathsf{t}_u \otimes \mathsf{t}_u\mathsf{S}_u^\top\widehat{\mathbf{q}_2}q_{2,3}^2\lambda^2, \quad \mathsf{t}_u = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ t_{2,2} & -t_{2,1} & 0 \end{pmatrix} \tag{17}$$

$$\widehat{\mathbf{L}_{2,(4,5,3)}} = \mathsf{D}_v^{-1}\mathsf{S}_v\mathsf{t}_v \otimes \mathsf{t}_v\mathsf{S}_v^\top\widehat{\mathbf{q}_2}(1-\lambda^2+q_{2,3}^2\lambda^2), \quad \mathsf{t}_v = \begin{pmatrix} 0 & t_{2,3} & -t_{2,2} \\ t_{2,3} & 0 & t_{2,1} \\ 0 & 0 & -1 \end{pmatrix} \tag{18}$$



(a) (b)

Figure 2: (a) shows an illustration of the refraction principle. (b) shows an image of "Snell's Window", which is the circle at the boundary of the image at which visibility of the sky ends. Photo courtesy gerb's photostream, http://www.flickr.com/photos/gerb/196296131/

Now, equation 15 can be modified to give the following

$$\begin{pmatrix}\widehat{\mathbf{q}_2} & \widehat{\mathbf{q}_2}q_{2,3}^2\end{pmatrix}\begin{pmatrix}0 & (1-\lambda^2)\Lambda_\mathsf{v}\\ \lambda^2\Lambda_\mathsf{u} & \lambda^2\Lambda_\mathsf{v}\end{pmatrix}\begin{pmatrix}(1-\lambda^2)\Lambda_\mathsf{s} & 0\\ \lambda^2\Lambda_\mathsf{s} & -\lambda^2\Lambda_\mathsf{t}\end{pmatrix}\bar{\mathsf{D}}_s\begin{pmatrix}\widehat{\mathbf{q}}\\ \widehat{\mathbf{q}}q_3^2\end{pmatrix} = 0 \tag{19}$$

$$\begin{pmatrix}\widehat{\mathbf{q}_2} & \widehat{\mathbf{q}_2}q_{2,3}^2\end{pmatrix}\underbrace{\begin{pmatrix}(1-\lambda^2)\lambda^2\Lambda_\mathsf{v}\Lambda_\mathsf{s} & -(1-\lambda^2)\lambda^2\Lambda_\mathsf{v}\Lambda_\mathsf{t}\\ (1-\lambda^2)\lambda^2\Lambda_\mathsf{u}\Lambda_\mathsf{s}+\lambda^4\Lambda_\mathsf{v}\Lambda_\mathsf{s} & -\lambda^4\Lambda_\mathsf{v}\Lambda_\mathsf{t}\end{pmatrix}\bar{\mathsf{D}}_s}_{\mathsf{F}}\begin{pmatrix}\widehat{\mathbf{q}}\\ \widehat{\mathbf{q}}q_3^2\end{pmatrix} = 0 \tag{20}$$

where $\Lambda_i = \mathsf{D}_i^{-1}\mathsf{S}_i\mathsf{t}_i \otimes \mathsf{t}_i\mathsf{S}_i^\top$.

Equation 20 defines the Fundamental matrix between two perspective cameras, when looking across a planar refractive surface. F is a matrix of dimensions 12×12 that relates the lifted coordinates in one image to a quartic curve in the other image. In a coordinate system where the refractive plane might be in a general position, this matrix is of dimensions 15×15. It is a function of the relative pose between the cameras, as well as the normal of the refractive plane and the refractive index. As expected, the quantity $\mathsf{F}\begin{pmatrix}\widehat{\mathbf{q}}\\ \widehat{\mathbf{q}}q_3^2\end{pmatrix}$ represents the epipolar quartic curve in the second image.

4 The Refractive Homography Matrix

In this section, we derive the relationship between a scene plane in 3D, and its image after refraction. Unlike the traditional case, we show that in the current scenario, this transformation is represented by a family of homographies that map one set of conics to another set.

Consider a cone of rays emerging from the camera, centered at the camera centre, with an axis that is aligned with the normal of the refractive plane and an aperture $2\theta_1$. Let us call this the incident cone (Figure 2a). Since any ray which is part of this cone surface, makes the same incident angle with the normal of the refractive plane, all outgoing rays from this cone make the same angle with the normal of the refractive plane. Thus the outgoing rays form a cone centered at the line

joining the camera centre and the plane, collinear with the normal of the refractive plane, with an aperture of $\theta_2 = 2\sin^{-1}\lambda\sin\theta_1$. Let us call this the refracted cone.

Since the image is formed by cutting the incident cone with the image plane, and the actual scene plane cuts the refracted cone, the 3D homography can be defined as a family of transformations, that transform one conic (formed by the scene plane and refracted cone) to another (formed by the image plane and the incident cone). This family is a function of θ_1.

Given a ray $\mathbf{q}$, the point $\mathbf{Q}$ on the refractive plane π and the refracted ray direction $\mathbf{R}$ are given by

$$\begin{aligned}
\mathbf{Q} &= \left(\begin{pmatrix}\mathbf{t}\\1\end{pmatrix}\mathbf{q}^\top - \mathbf{q}\begin{pmatrix}\mathbf{t}^\top & 1\end{pmatrix}\right)\pi && (21)\\
&= \begin{pmatrix}(\tilde{\mathbf{q}}\times\mathbf{t})_\times & -\tilde{\mathbf{q}}\\ \tilde{\mathbf{q}}^\top & 0\end{pmatrix}\pi && (22)\\
&= \begin{pmatrix}(\tilde{\mathbf{q}}\times\mathbf{t})_\times\mathbf{v} - \tilde{\mathbf{q}}d\\ \cos(\theta_1)\end{pmatrix} = \begin{pmatrix}(\mathbf{v})_\times(\mathbf{t})_\times\tilde{\mathbf{q}} - \tilde{\mathbf{q}}d\\ \cos(\theta_1)\end{pmatrix} && (23)\\
\mathbf{R} &= \begin{pmatrix}q_1\lambda & q_2\lambda & \cos(\theta_2) & 0\end{pmatrix}^\top && (24)
\end{aligned}$$

where θ_1 is the incident angle and θ_2 is the refracted angle, and $(\cdot)_\times$ represents a skew symmetric matrix that corresponds to a cross product with the vector inside parenthesis. The point $\mathbf{S}$ where the above line will strike a plane $\pi_1 = \begin{pmatrix}\mathbf{v}_1^\top & d_1\end{pmatrix}$ is now given by

$$\begin{aligned}
\mathbf{S} &= \left(\mathbf{R}\mathbf{Q}^\top - \mathbf{Q}\mathbf{R}^\top\right)\pi_1 && (25)\\
&= \begin{pmatrix}(\mathbf{Q}_{(1,2,3)}\times\mathbf{R}_{(1,2,3)})_\times & \cos(\theta_1)\mathbf{R}_{(1,2,3)}\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)} & 0\end{pmatrix}\pi_1 && (26)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{Q}_{(1,2,3)}\times\mathbf{R}_{(1,2,3)}) + \cos(\theta_1)\mathbf{R}_{(1,2,3)}d_1\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)}\cdot\mathbf{v}_1\end{pmatrix} && (27)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)} + (\mathbf{v}_1)_\times(\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)})d + \cos(\theta_1)\mathbf{R}_{(1,2,3)}d_1\\ -\cos(\theta_1)\mathbf{R}_{(1,2,3)}\cdot\mathbf{v}_1\end{pmatrix} && (28)\\
&= \begin{pmatrix}-(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\left(\mathbf{v}_1\right)_\times & d_1\mathbf{I}_{3\times3}\\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)}\\ \cos(\theta_1)\mathbf{R}_{(1,2,3)}\end{pmatrix} && (29)
\end{aligned}$$

The vector $\tilde{\mathbf{q}}\times\mathbf{R}_{(1,2,3)}$ expands to give $\begin{pmatrix}q_2(\cos(\theta_2)-\cos(\theta_1))\\ q_1(\cos(\theta_1)-\cos(\theta_2))\\ 0\end{pmatrix}$. We can isolate the cosines into

a separate matrix to get

$$\mathbf{S} = \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\left(\mathbf{v}_1\right)_\times & d_1\mathsf{I}_{3\times3} \\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top \end{pmatrix} \begin{pmatrix} q_2(\cos(\theta_2)-\cos(\theta_1)) \\ q_1(\cos(\theta_1)-\cos(\theta_2)) \\ 0 \\ \cos(\theta_1)q_1\lambda \\ \cos(\theta_1)q_2\lambda \\ \cos(\theta_1)\cos(\theta_2) \end{pmatrix} \tag{30}$$

$$= \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times + d\left(\mathbf{v}_1\right)_\times & d_1\mathsf{I}_{3\times3} \\ \begin{pmatrix}0 & 0 & 0\end{pmatrix} & -\mathbf{v}_1^\top \end{pmatrix} \begin{pmatrix} 0 & a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \\ \lambda\cos(\theta_1) & 0 & 0 \\ 0 & \lambda\cos(\theta_1) & 0 \\ 0 & 0 & \cos(\theta_2) \end{pmatrix} \tilde{\mathbf{q}} \tag{31}$$

$$= \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\left(\mathbf{a}\right)_\times + d\left(\mathbf{v}_1\right)_\times\left(\mathbf{a}\right)_\times + d_1\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \\ -\mathbf{v}_1^\top\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \end{pmatrix} \tilde{\mathbf{q}} \tag{32}$$

$$\tag{33}$$

$$\sim \mathbf{H}_{\theta_1}\mathbf{x} \tag{34}$$

with

$$\mathbf{H}_{\theta_1} = \begin{pmatrix} -(\mathbf{v}_1)_\times(\mathbf{v})_\times(\mathbf{t})_\times\left(\mathbf{a}\right)_\times + d\left(\mathbf{v}_1\right)_\times\left(\mathbf{a}\right)_\times + d_1\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \\ -\mathbf{v}_1^\top\begin{pmatrix}\lambda\cos(\theta_1)\mathsf{I}_{2\times2} & \mathbf{0} \\ \mathbf{0}^\top & \cos(\theta_2)\end{pmatrix} \end{pmatrix} \mathsf{R}^\top \tag{35}$$

$$a = \cos(\theta_2) - \cos(\theta_1) \tag{36}$$

$$\mathbf{a} = \begin{pmatrix}0 & 0 & a\end{pmatrix}^\top \tag{37}$$

Equation 35 shows how the homography matrix can be expressed as a function of the incident and refracted angles. From the above expression, it is clear that there exists a *family* of homographies that are linear in the cosines of these angles, which fully represent the projection of a plane onto the image. Although the above expression can be further developed to obtain an expression in lifted coordinates that is independent of these angles, i.e., a homography that applies to all points of the image and scene plane, we omit the derivation here because of space constraints.

5 Exploiting Snell's Window

Till now, we have discussed the general case of planar refraction, which is applicable to any scene irrespective of the relative refractive index. Of particular importance, is a special case, when the camera is in a denser medium (say water) and looking outward into a lighter medium (say air). This amounts to a relative refractive index that is > 1 (as per our current convention). In this case, it is possible to image the *peripheral rays*: rays that make an angle of $\pi/2$ with the refractive surface

normal after refraction, and hence are parallel to the refracting surface. The importance of these rays is that any ray with an incident angle greater than the incident angle of these rays is *reflected*. This set of rays thus captures the periphery of the world on the other side of the surface, which in the case of underwater imagery is the horizon (Figure 2b).

5.1 The Image of Snell's Window

Consider a point $\mathbf{x}$ that is on this periphery. Since after refraction the ray from this point is parallel to the refractive surface, the refracted angle is $\frac{\pi}{2}$, which means

$$1 - \lambda^2 + \lambda^2(\pi\mathbf{q})^2 = 0 \tag{38}$$

where $\mathbf{q} = \begin{pmatrix}(\mathsf{R}^\top\hat{\mathbf{x}})^\top & 0\end{pmatrix}^\top$, with $\tilde{\mathbf{x}}$ being the unit vector of image point $\mathbf{x}$ in the camera coordinate system. Let us develop these equations further.

$$\begin{aligned}
\mathbf{q}^\top\pi\pi^\top\mathbf{q} &= 1 - \frac{1}{\lambda^2} && (39)\\
\tilde{\mathbf{x}}^\top\mathsf{R}^\top\mathbf{v}\mathbf{v}^\top\mathsf{R}\tilde{\mathbf{x}} &= 1 - \frac{1}{\lambda^2} && (40)\\
\frac{1}{\|\mathbf{x}^2\|}\mathbf{x}^\top\mathsf{R}^\top\mathbf{v}\mathbf{v}^\top\mathsf{R}\mathbf{x} &= 1 - \frac{1}{\lambda^2} && (41)\\
\mathbf{x}^\top\mathsf{R}^\top\mathbf{v}\mathbf{v}^\top\mathsf{R}\mathbf{x} &= \mathbf{x}^\top\mathsf{I}_{3\times3}\mathbf{x}(1 - \frac{1}{\lambda^2}) && (42)\\
\mathbf{x}^\top\left(\mathsf{R}^\top\mathbf{v}\mathbf{v}^\top\mathsf{R} - (1 - \tfrac{1}{\lambda^2})\mathsf{I}_{3\times3}\right)\mathbf{x} &= 0 && (43)
\end{aligned}$$

which shows that the periphery is a conic in image coordinates. The term $\mathsf{R}^\top\mathsf{v}$ represents the refractive plane normal in the camera coordinate system. One of the main advantages of Equation 43 is that one of the conics's eigenvectors is the normal of the refractive plane.

$$\left(\mathsf{R}^\top\mathbf{v}\mathbf{v}^\top\mathsf{R} - (1 - \tfrac{1}{\lambda^2})\mathsf{I}_{3\times3}\right)\left(\mathsf{R}^\top\mathbf{v}\right) = \frac{1}{\lambda^2}\mathsf{R}^\top\mathsf{v} \tag{44}$$

since $\mathsf{v}^\top\mathsf{v} = 1$.

In fact, it is easy to compute that the eigenvalues of the above matrix are $\frac{1}{\lambda^2}, -1 + \frac{1}{\lambda^2}, -1 + \frac{1}{\lambda^2}$. Thus the important property of the eigenvalues and eigenvectors of the above matrix is that *the only positive eigenvalue is the square inverse of the relative refractive index, while the corresponding eigenvector is the normal of the refractive plane in the camera coordinate system.* Thus we have shown that the image of the horizon after refraction from a planar scene is a conic, and that using a simple SVD based algorithm it is possible to recover both the plane normal and the relative refractive index from a *single* image.

5.2 Relative Pose Computation

Given that we can estimate the plane normal (in the camera coordinate system), and the refractive index, we now show how to compute the relative pose of a second camera with respect to the first. Let us first note that since both the cameras see the same refractive surface, estimating the normal of the plane in the camera coordinate systems of both cameras can help estimating the rotation

between them upto a rotation *about* the normal itself. Additionally, we can set the perpendicular distance from the camera centre to the refractive plane to 1, so as to set the scale of the scene. Since we have chosen a world coordinate system where the plane normal is aligned with the Z-axis, there are 4 unknowns that remain to be computed, namely the rotation about the Z-axis and the translation parameters.

In order to solve for these parameters, let us start by observing the fundamental matrix equation 20. Observe that the image points (rays) are a function of the unknown rotation matrix of the second camera, which is of the form $\mathsf{R} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Thus, the vector $\begin{pmatrix} \widehat{\mathbf{q}_2} & \widehat{\mathbf{q}_2} q_{2,3}^2 \end{pmatrix}^\top$ is quadratic in the elements of the rotation matrix R. The fundamental matrix F on the other hand, is only a function of the translation of the camera, and so its elements are quadratic in the elements of translation. Thus, 4 point-correspondences are sufficient to solve for these biquadratic equations.

6 Discussion and Conclusion

In this paper, we have defined for the first time the epipolar geometry for scenes where a single plane separates two media with different refractive indices. The camera(s) are on one media, while the object(s) being observed are on the other side. Such scenarios occur often in underwater vision. We have shown that for the case when the camera is in a denser medium, it is even possible to estimate the refractive index and the refractive surface geometry from a *single* image, by observing the so-called "Snell's Window". Further, we have shown a method to compute the relative pose between two cameras in such a scenario.

The contributions of this work are theoretical, and of conceptual value. Our main motivation for this work has been to explicitly model the geometry behind imaging in such scenarios. Apart from the theoretical value of the results we have shown thus far, what is also important is that using the results of this paper, it is possible to *measure* the deviation of the standard perspective model in areas like underwater imagery. This is of importance while modeling errors in observation using a perspective model.

References

[1] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? In *ICCV 2003*, volume 2, pages 1025–1032, Nice, France, 2003.

[2] G. Glaeser and H.-P. Schröcker. Reflections on refractions. *Journal for Geometry and Graphics*, 4(1):1–18, 2000.

[3] N. Gracias and J. Santos-Victor. Underwater video mosaics as visual navigation maps. *Computer Vision and Image Understanding*, 79(1):66–91, 2000.

[4] R. Horn and C. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991.

[5] K. N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *ICCV 2005*, volume 2, pages 1448–1455, Beijing, China, 2005.

[6] J.-M. Lavest, G. Rives, and J.-T. Lapresté. Underwater camera calibration. In *Proceedings of ECCV 2000-Part II*, pages 654–668, Dublin, Ireland, 2000.

[7] H.-G. Maas. New developments in multimedia photogrammetry. In *Optical 3D Measurement Techniques*, Kalsruhe, 1995. Wichmann Verlag.

[8] N. J. W. Morris and K. N. Kutulakos. Dynamic refraction stereo. In *ICCV 2005*, volume 2, pages 1573–1580, Beijing, China, 2005.

[9] S. Negahdaripour, H. Sekkati, and H. Pirsiavash. Opit-acoustic stereo imaging, system calibration and 3-d reconstruction. In *CVPR 2007: Beyond Multiview Geometry Workshop*, pages 1–8, Minneapolis, USA, 2007.

[10] O. Pizarro, R. Eustice, and H. Singh. Relative pose estimation for instrumented, calibrated imaging platforms. In *DICTA 2003*, pages 601–612, Sydney, Australia, 2003.

[11] H. Pottmann and J. Wallner. *Computational line geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

[12] M. R. Shortis and E. S. Harvey. Design and calibration of an underwater stereo-vision system for the monitoring of marine fauna populations. *International Archives of Photogrammetry and Remote Sensing*, 32:792–799, 1998.

[13] P. Sturm and J. P. Barreto. General imaging geometry for central catadioptric cameras. In *ECCV 2008-Part II*, volume 4, pages 609–622, Marseille, France, 2008.

[14] T. Svoboda and T. Pajdla. Epipolar geometry for central catadioptric cameras. *International Jounral of Computer Vision*, 49(1):23–37, 2002.

[15] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. In *CVPR 2008*, pages 1–8, Los Alamitos, CA, USA, 2008.