A Theory of Multi-Layer Flat Refractive Geometry

Amit Agrawal, Srikumar Ramalingam, Yuichi Taguchi
Mitsubishi Electric Research Labs (MERL)
[agrawal,ramalingam,taguchi] at merl.com

Visesh Chari
INRIA Rhône-Alpes
visesh.chari at inrialpes.fr

Abstract

Flat refractive geometry corresponds to a perspective camera looking through single/multiple parallel flat refractive mediums. We show that the underlying geometry of rays corresponds to an axial camera. This realization, while missing from previous works, leads us to develop a general theory of calibrating such systems using 2D-3D correspondences. The pose of 3D points is assumed to be unknown and is also recovered. Calibration can be done even using a single image of a plane.

We show that the unknown orientation of the refracting layers corresponds to the underlying axis, and can be obtained independently of the number of layers, their distances from the camera and their refractive indices. Interestingly, the axis estimation can be mapped to the classical essential matrix computation and 5-point algorithm [15] can be used. After computing the axis, the thicknesses of layers can be obtained linearly when refractive indices are known, and we derive analytical solutions when they are unknown. We also derive the analytical forward projection (AFP) equations to compute the projection of a 3D point via multiple flat refractions, which allows non-linear refinement by minimizing the reprojection error. For two refractions, AFP is either 4^{th} or 12^{th} degree equation depending on the refractive indices. We analyze ambiguities due to small field of view, stability under noise, and show how a two layer system can be well approximated as a single layer system. Real experiments using a water tank validate our theory.

1. Introduction

A camera observing a scene through multiple refractive planes (*e.g.* underwater imaging) results in distortions and gives the illusion of scene being closer and magnified. While 3D reconstruction in such scenarios has been analyzed in multi-media photogrammetry [9, 20, 18], such imaging setups have been relatively unaddressed in computer vision community until recently. Calibrating such a system with multiple layers with unknown layer orientation, distances and refractive indices remains an open and challenging problem.

The fact that such systems do not correspond to a single viewpoint system is known (see, for example, [25]). However, we show that the underlying geometry of rays in such systems actually corresponds to an axial camera. This realization, which has been missing from previous works to the best of our knowledge, allows us to handle multiple layers in a unified way and results in practical and robust algorithms. Firstly, we show that the unknown orientation of the refractive layers corresponds to the underlying axis, which can be estimated *independently* of the number of layers, their distances and their refractive indices. This results in considerable simplification of the calibration problem via a two-step process, where the axis is computed first. Without such simplification, the calibration is difficult to achieve. Secondly, we show that the axis estimation can be mapped to the classical relative orientation problem (essential matrix estimation) for which excellent solutions (*e.g.* 5-point algorithm [15]) already exist. In fact, calibration can be done using a single plane similar to [23]. Our primary contributions are as follows.

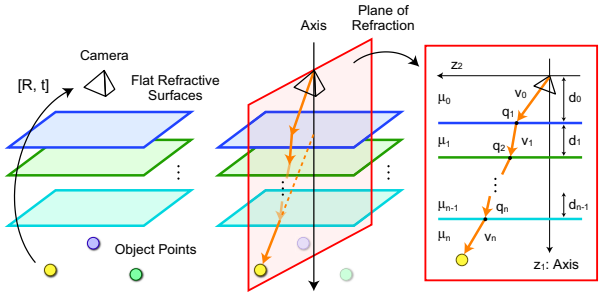


Figure 1. (Left) Flat refractive geometry with n layers. (Middle) The entire light-path for each pixel lies on a plane and all planes intersect in a common axis passing through the camera center. (Right) After computing the axis, analysis can be done on the plane of refraction to estimate layer thicknesses and refractive indices.

- We show that the geometry of rays in flat refraction systems corresponds to an axial camera, leading to a unified theory for calibrating such systems with multiple layers.
- By demonstrating the equivalence with classical essential matrix estimation, we propose efficient and robust algorithms for calibration using planar as well as non-planar objects.

- We derive analytical forward projection equations for multiple refractions, which allows minimizing the reprojection error.
- We analyze ambiguities with small FOV, and show that multiple layer systems may be well-approximated by single/two-layer systems.

1.1. Related Work

Maas [13] considered a three layer system assuming that the image plane is parallel to the refractive interfaces. His approach corrects for the *radial* shift of the projected 3D points using optimization. Treibitz *et al*. [25] considered a single refraction with known refractive index in an underwater imaging scenario. They assume the distance of the interface as the single unknown parameter (when the camera is internally calibrated) and perform calibration using known depth of a planar checkerboard. The image plane is parallel to the interface in their setup as well. In contrast, we (a) do not assume that the refractive interfaces are fronto-parallel, (b) handle multiple layers with unknown layer distances, (c) consider known/unknown refractive indices, and (d) do not assume known pose of the calibrating object. We only assume that the camera is internally calibrated.

3D reconstruction under reflections/refractions has been explored in [3, 14, 19, 1, 21] either for reconstructing the scene or the medium itself. Chen *et al*. [3] captured two images, with and without a thick glass slab for 3D reconstruction. Both images are required to estimate the orientation of the slab and an additional image to obtain the refractive index. We show that a single set of 2D-3D correspondences from a single photo allows estimating medium thickness as well as refractive index. Other works assume known vertical direction [1] or require several images for calibration [19]. Steger and Kutulakos [21] showed that light-path triangulation becomes degenerate when the entire light-path lies on a plane, which is the case here. Their goal is to compute the shape of the refractive medium, and they consider each light-path independently. In contrast, we know partial knowledge of shape (flat parallel layers), and light-paths can be *parameterized*. Thus, we can use information from multiple-light paths to compute the calibration parameters. For two refractions (air-medium-air), our analysis is consistent with [21] in that the distance to the medium cannot be estimated. However, we show theoretically that if all refractive indices are different, light paths are not degenerate for any number of layers.

Refractions have also been modeled using ray-tracing [10, 12] for calibration. Kotowski [10] proposed a *bundle-triangulation* framework where the points of refraction are computed iteratively, starting with a central approximation. In contrast, we derive analytical forward projection (AFP) equations for computing the projection of a 3D point via multiple refractions. This allows non-linear refinement of the initial solution by minimizing the reprojection error. For single refraction, AFP is a 4^{th} degree equation [5]. For two refractions, we derive a $4^{th}/12^{th}$ degree equation depending on refractive indices.

Non-Central/Axial Cameras: Pless [16] proposed algorithms for relative motion estimation for calibrated non-central cameras. Chari and Sturm [2] showed the existence of geometric entities such as fundamental matrix considering refraction. Generic camera calibration algorithms [22] have been proposed to calibrate non-central cameras. Li *et al*. [11] analyzed the degeneracies in axial cameras for motion estimation. Ramalingam *et al*. [17] proposed a general framework for calibrating axial cameras using three checkerboards. Their parameterization does not use camera-side rays and involves two rotations/translations. In contrast, we parameterize the axis to pass through the camera and use camera-side rays, allowing calibration from a single plane. However, we use their insight that the axis can be computed first. Models based on radial distortion for calibration assume known center of distortion [24] or model each distortion circle separately [23]. We use a global model with fewer parameters for flat refractive systems.

2. General Flat Refractive Geometry

Consider the general setup for flat refractive geometry as shown in Figure 1, where a perspective camera observes a known calibration object via n flat refraction layers. We work in the camera coordinate system with the camera center at the origin. Let $\mathbf{n}$ denote the common surface normal to all the layers and $[d_i, \mu_i]$ be the thickness and refractive index of the i^{th} medium respectively. d_0 represents the distance between the camera and the first layer. Let $\mathbf{P}(i)_{i=1}^{K}$ denote K 3D points on the object which are known in the object coordinate system and let $[R, \mathbf{t}]$ be the unknown rigid transformation of these points.

2.1. Flat Refraction Constraint

We assume that the internal camera calibration has been done offline and hence we know the camera ray $\mathbf{v}_0(i)$ for each 3D point $\mathbf{P}(i)$. Let $[\mathbf{v}_0(i), \mathbf{v}_1(i), \ldots, \mathbf{v}_n(i)]$ denote the direction vectors of each segment of the corresponding light-path. The last refracted ray $\mathbf{v}_n$ should be parallel to the line joining the transformed 3D point $R\mathbf{P}+\mathbf{t}$ and the refraction point $\mathbf{q}_n$ on the last layer. Thus, the following Flat Refraction Constraint (FRC) should be satisfied.

$$\texttt{FRC}: \quad (R\mathbf{P} + \mathbf{t} - \mathbf{q}_n) \times \mathbf{v}_n = 0, \tag{1}$$

where $\times$ denotes the cross-product. Our goal is to estimate the unknown calibration parameters $\mathbf{n}, [d_i]_{i=0}^{n-1}, [\mu_i]_{i=0}^{n}$ as well as the unknown pose $[R, \mathbf{t}]$ given K 2D-3D correspondences $[\mathbf{v}_0(i), \mathbf{P}(i)]_{i=1}^{K}$.

From Snell's law, $\mu_i \sin\theta_i = \mu_{i+1} \sin\theta_{i+1}$, where θ_i is the angle between $\mathbf{v}_i$ and $\mathbf{n}$. This can be written in vector

form [6] as

$$\mathbf{v}_{i+1} = a_{i+1}\mathbf{v}_i + b_{i+1}\mathbf{n}, \quad (2)$$

where $a_{i+1} = \mu_i/\mu_{i+1}$ and

$$b_{i+1} = \frac{-\mu_i \mathbf{v}_i^T\mathbf{n} - \sqrt{\mu_i^2(\mathbf{v}_i^T\mathbf{n})^2 - (\mu_i^2 - \mu_{i+1}^2)\mathbf{v}_i^T\mathbf{v}_i}}{\mu_{i+1}}. \quad (3)$$

Since Snell's law only depends on the ratio of the refractive indices, we assume $\mu_0 = 1$ without loss of generality. We first derive the FRC for a single layer and a single 3D point $\mathbf{P}$. The refraction point $\mathbf{q}_1$ equals $-d_0\mathbf{v}_0/(\mathbf{v}_0^T\mathbf{n})$. Substituting in (1),

$$(R\mathbf{P}+\mathbf{t}) \times (a_1\mathbf{v}_0 + b_1\mathbf{n}) + b_1 d_0(\mathbf{v}_0 \times \mathbf{n})/(\mathbf{v}_0^T\mathbf{n}) = 0. \quad (4)$$

After substituting for a_1 and b_1 and removing the square root term in b_1, we get an equation with second order terms of R, $\mathbf{t}$ and μ_i and sixth order terms of $\mathbf{n}$. Thus, directly solving the FRC is quite difficult. More importantly, the complexity of the FRC equation increases with each additional layer due to the square root term in each b_i. Thus, in order to solve this problem efficiently, we need to analyze the geometry of underlying rays to derive simpler constraints. In the next section, we describe coplanarity constraints that allow us to estimate the normal $\mathbf{n}$ and 5 out of 6 pose parameters *independently* of d_i's and μ_i's.

3. Coplanarity Constraints

Axial Camera: We first show that an n-layer flat refraction system corresponds to an axial camera. The axis is defined as the line parallel to $\mathbf{n}$ passing through the camera center (origin) and let $\mathbf{A}$ be its direction vector. Let π be the plane of refraction (POR) containing the axis and a given camera ray. The normal $\mathbf{n}$ lies on π. From Snell's law, the incoming ray, the normal and the refracted ray lie on the same plane at any refraction boundary. Hence, by induction, the entire light-path should lie on π and the last refracted ray should intersect the axis, since both of them are coplanar. Thus, all outgoing rays intersect the axis and the system is axial.

The transformed 3D point $R\mathbf{P}+\mathbf{t}$ should also lie on π. Thus, the coplanarity constraint for each 3D point can be written as

$$\texttt{Coplanarity}: \quad (R\mathbf{P}+\mathbf{t})^T(\mathbf{A} \times \mathbf{v}_0) = 0, \quad (5)$$

where $(\mathbf{A} \times \mathbf{v}_0)$ is the normal to POR. Note that the coplanarity constraint is *independent* of the number of layers n, their thicknesses d_i, and the refractive indices μ_i. It only depends on the axis and pose parameters.

3.1. Axis Computation

Let $X(:)$ be the vector formed by stacking the columns of a matrix X and let $\otimes$ denote the kronecker product. Let $[\mathbf{A}]_\times$ be the 3×3 skew-symmetric matrix obtained from 3-vector $\mathbf{A}$. The coplanarity constraint can be re-written as

$$0 = \mathbf{v}_0^T(\mathbf{A} \times (R\mathbf{P}+\mathbf{t})) = \mathbf{v}_0^T E\mathbf{P} + \mathbf{v}_0^T\mathbf{s}, \quad (6)$$

where $E = [\mathbf{A}]_\times R$ and $\mathbf{s} = \mathbf{A} \times \mathbf{t}$. Note that $\mathbf{s}^T\mathbf{A} = 0$ and thus the full translation $\mathbf{t}$ cannot be estimated using coplanarity constraints. The component of $\mathbf{t}$ in the direction of axis, $\mathbf{t}_A$, vanishes in $\mathbf{s}$. Thus, we have 7 degrees of freedom: 2 for axis, 3 for rotation and 2 for translation.

11-point Linear Algorithm: Stacking equations for 11 correspondences, we get a linear system

$$\underbrace{\begin{bmatrix} (\mathbf{P}(1)^T \otimes \mathbf{v}_0(1)^T) & \mathbf{v}_0(1)^T \\ \vdots & \vdots \\ (\mathbf{P}(11)^T \otimes \mathbf{v}_0(11)^T) & \mathbf{v}_0(11)^T \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} E(:) \\ \mathbf{s} \end{bmatrix} = 0, \quad (7)$$

where B is a 11×12 matrix whose rank is 11. Let $B = U\Sigma V^T$ be the SVD of B. The solution is given by the right null singular vector of B (last column of V). The scale factor is obtained by setting the norm of E to one.

8-point Algorithm: Notice the striking similarity between our E matrix ($[\mathbf{A}]_\times R$) and the essential matrix [8] for relative motion between two perspective cameras ($[\mathbf{t}]_\times R$). This implies that we can map the axis estimation to the 5-point algorithm for essential matrix computation [15]. Given 8 correspondences, we obtain a 8×12 matrix B as above. Let $\mathbf{V}_{i=1}^4$ be the right null singular vectors of B. The solution lies in a four dimensional subspace

$$\begin{bmatrix} E(:) \\ \mathbf{s} \end{bmatrix} = \lambda_1\mathbf{V}_1 + \lambda_2\mathbf{V}_2 + \lambda_3\mathbf{V}_3 + \lambda_4\mathbf{V}_4, \quad (8)$$

where λ_i's are unknown scalars. λ_4 can be set to 1 due to the unknown scale factor. The 'E' part of the solution is

$$E = \lambda_1\mathbf{V}_1(1:9) + \lambda_2\mathbf{V}_2(1:9) + \lambda_3\mathbf{V}_3(1:9) + \mathbf{V}_4(1:9),$$

where $\mathbf{V}_i(1:9)$ denote the first 9 elements of $\mathbf{V}_i$. Now λ_i's can be computed using the solution in [15] by providing the above subspace vectors for E.

After recovering E and $\mathbf{s}$, the axis is computed as the left null singular vector of E (since $\mathbf{A}^T E = 0$). The sign ambiguity in axis is resolved by pointing it away from the camera. The translation orthogonal to axis, $\mathbf{t}_{A^\perp}$, can be obtained as $\mathbf{s} \times \mathbf{A}$. Four solutions for R are recovered from E as in [15]. The correct solution is obtained after recovering the layer thicknesses and constraining them to be positive. Interestingly, the axis estimation is similar to the center of distortion estimation for central cameras in [7] and our 8pt algorithm can be applied.

3.2. Simulations

We present simulations for estimating the axis with Gaussian noise (variance σ^2 pixels) in feature points for

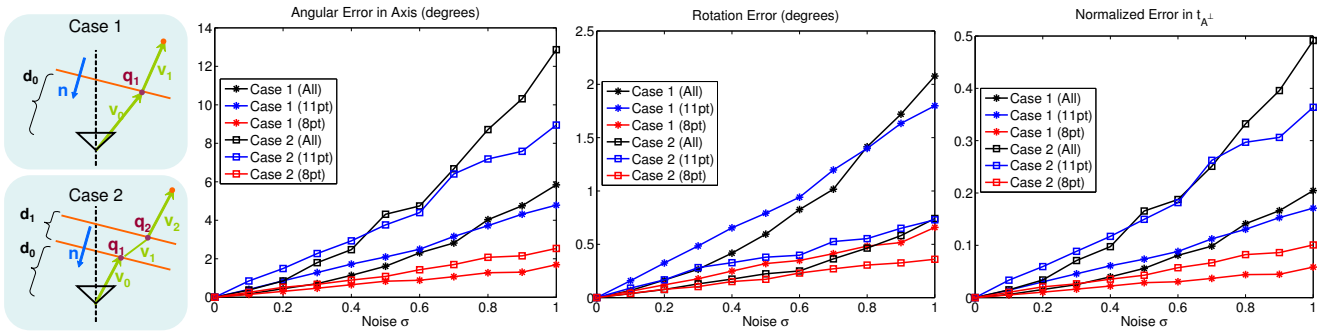


Figure 2. (Left) Single layer (Case 1) and two layer (Case 2, $\mu_2 = \mu_0$) configurations. (Right) Comparison of 11pt algorithm, 8pt algorithm and using all points in a least squares fashion for estimation of axis, rotation and $\mathbf{t}_{A^\perp}$ using coplanarity constraints.

Case 1 and Case 2, shown in Figure 2. We assume a camera with FOV of 45° and resolution 1000^2 pixels. The scene is set by choosing $d_0 = 300$ units, $d_1 = 450$ units, $\mu_1 = 1.5$ and $\mu_2 = 1$. We perform 100 trials for each noise setting, and plot the average error in axis, rotation and $\mathbf{t}_{A^\perp}$ in Figure 2. For each trial, the axis is randomly generated in a cone of half-angle 45° around the camera's optical axis. 100 3D points are randomly generated along with R and $\mathbf{t}$ so that they lie within $[d_0, 2d_0]$ after the last layer. For each trial, RANSAC-framework is employed for both 11pt and 8pt algorithms using 200 iterations to choose the best solution using the coplanarity error. Rotation error is defined as the minimum angle of rotation required to go from the estimated rotation to the true rotation. Similarly, axis error is defined as the angle between the estimated and the true axis. The translation error is computed as the norm of the translation error vector and is normalized using the corresponding layer thickness. Notice that when $\mathbf{v}_0 \parallel \mathbf{n}$, the coplanarity constraint cannot be used. Thus, rays having small angle with axis are unreliable. As expected, the 8pt algorithm performs significantly better than the 11pt algorithm as well as using all points in a least squares fashion.

4. Recovery of Layer Thicknesses

We first assume known refractive indices. Section 3.1 showed how to compute the axis $\mathbf{A}$, rotation R and translation $\mathbf{t}_{A^\perp}$ orthogonal to the axis. When μ_i's are known, the ray *directions* for the entire light-path $\mathbf{v}_0(i), \ldots, \mathbf{v}_n(i)$ can be pre-computed using the estimated $\mathbf{A}$. The remaining unknowns are the layer thickness d_i's and the translation $\mathbf{t}_A$ along the axis, which can be computed *linearly* as described below.

Coordinate Transformations: We first apply the computed R and $\mathbf{t}_{A^\perp}$ to the 3D points $\mathbf{P}$. Let $\mathbf{P}_c = R\mathbf{P} + \mathbf{t}_{A^\perp}$. With known axis, the analysis can be done in 2D on the plane of refraction (POR) itself as shown in Figure 1. Let $\mathbf{t}_A = \alpha\mathbf{A}$, where α is the unknown translation magnitude along the axis. Let $[\mathbf{z}_2, \mathbf{z}_1]$ denote an orthogonal coordinate system on the POR. We choose $\mathbf{z}_1$ along the axis. For a given camera ray $\mathbf{v}_0$, let $\mathbf{z}_2 = \mathbf{z}_1 \times (\mathbf{z}_1 \times \mathbf{v}_0)$ be the orthogonal direction. The projection of $\mathbf{P}_c$ on POR is given by $\mathbf{u} = [u^x, u^y]$, where $u^x = \mathbf{z}_2^T\mathbf{P}_c$ and $u^y = \mathbf{z}_1^T\mathbf{P}_c$. Similarly, the direction vector $\mathbf{v}_i$ of each ray on the light-path of $\mathbf{v}_0$ can be represented by a 2D vector $\mathbf{vp}_i$ on POR, whose components are given by $\mathbf{z}_2^T\mathbf{v}_i$ and $\mathbf{z}_1^T\mathbf{v}_i$. Let $c_i = \mathbf{vp}_i^T\mathbf{z}_1$ and $\mathbf{z}_p = [0; 1]$ be a unit 2D vector.

4.1. Linear System for n Layers

For each correspondence, the FRC for n layer system on its plane of refraction is given by

$$\mathbf{vp}_n \times (\mathbf{u} + \alpha\mathbf{z}_p - \mathbf{q}_n) = 0. \tag{9}$$

This is because the last refracted ray $\mathbf{vp}_n$ should be parallel to the line joining the transformed 3D point $\mathbf{u} + \alpha\mathbf{z}_p$ and the refraction point $\mathbf{q}_n$ on the last layer. $\mathbf{q}_n = \sum_{i=0}^{n-1} -d_i\mathbf{vp}_i/c_i$. Substituting, we get

$$\mathbf{vp}_n \times \begin{bmatrix} \frac{\mathbf{vp}_0}{c_0} & \ldots & \frac{\mathbf{vp}_{n-1}}{c_{n-1}} & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ \vdots \\ d_{n-1} \\ \alpha \end{bmatrix} = -\mathbf{vp}_n \times \mathbf{u}.$$

Thus, each correspondence gives one linear equation in d_i's and α. By stacking $K > n$ correspondences, the resulting linear system can be solved to obtain d_i's and α for n layers. After estimating α, the translation $\mathbf{t}$ is given by $\mathbf{t}_{A^\perp} + \alpha\mathbf{A}$. However, if $\mu_i = \mu_n$ for any i, $\mathbf{vp}_i \parallel \mathbf{vp}_n$ and d_i cannot be estimated. In addition, if $\mu_i = \mu_j$, only the *combined* layer thickness $d_i + d_j$ can be estimated, since the corresponding columns in the linear system become equal. Now we analyze some special cases.

Case 1 **(Single Refraction):** For single layer, we have two unknowns d_0 and α and the FRC is given by

$$\mathbf{vp}_1 \times \begin{bmatrix} \mathbf{vp}_0/c_0 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ \alpha \end{bmatrix} = -\mathbf{vp}_1 \times \mathbf{u}. \tag{10}$$

Using $K \geq 2$ correspondences, a least squares solution can be obtained.

Case 2 **(Two Refractions)** $\mu_0 = \mu_2$**:** This is a common scenario when looking *through* a refractive medium such as

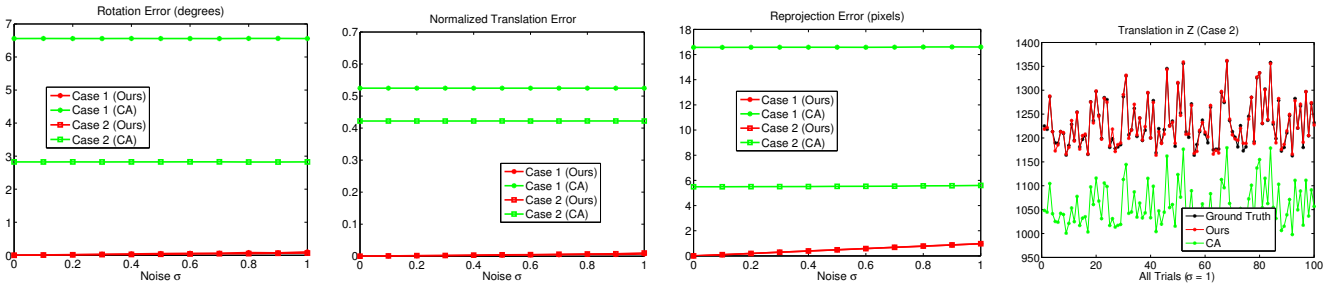


Figure 3. Rotation, translation and reprojection error using our algorithm versus using a central approximation (CA) for Case 1 and Case 2. The right most plot shows the estimated $\mathbf{t}_z$ for Case 2 over all 100 trials for $\sigma = 1$ pixel. CA estimates the object to be closer to the camera.

a thick glass slab. Here d_0, d_1 and α are unknowns. Since $\mu_0 = \mu_2$, $\mathbf{vp}_2 \parallel \mathbf{vp}_0$ and the FRC reduces to

$$\mathbf{vp}_0 \times \begin{bmatrix} \mathbf{vp}_1/c_1 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_1 \\ \alpha \end{bmatrix} = -\mathbf{vp}_0 \times \mathbf{u}. \quad (11)$$

Thus, we can only estimate the thickness d_1 of the medium, but not its distance d_0. This is consistent with the analysis shown in [21]. In Section 5, we show that although d_0 cannot be estimated, the projection of the 3D point on the image plane can be computed for non-linear refinement.

Case 3 (Two Refractions) $\mu_0 \neq \mu_2$**:** Now $\mathbf{v}_2$ and $\mathbf{v}_0$ are not parallel and the FRC is given by

$$\mathbf{vp}_2 \times \begin{bmatrix} \mathbf{vp}_0/c_0 & \mathbf{vp}_1/c_1 & \mathbf{z}_p \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \\ \alpha \end{bmatrix} = -\mathbf{vp}_2 \times \mathbf{u}. \quad (12)$$

Thus, we can estimate the distance d_0 as well.

5. Non-Linear Refinement

Till now we described an initial solution for estimating the calibration and pose parameters. We now show how to perform non-linear refinement of calibration and pose parameters.

Analytical Forward Projection (AFP): Given a *calibrated* central or non-central camera, the AFP describes an analytical method to compute the projection (or the corresponding camera ray) of a known 3D point. AFP can be used to minimize the image reprojection error. For Case 1, AFP is a 4^{th} degree equation [5]. We derive the AFP equation for two refractions and show that it is a 4^{th} degree equation for Case 2 and a 12^{th} degree equation for Case 3. The analysis can be done on POR.

Case 2: Assume given calibration parameters d_0, d_1, μ_1 and the coordinates $\mathbf{u} = [u^x, u^y]$ of a known 3D point on POR. The unknown camera ray $\mathbf{vp}_0$ can be parameterized as $[x, d_0]$. Since $\mu_2 = \mu_0$, $\mathbf{vp}_2 \parallel \mathbf{vp}_0$. On POR, the normal $\mathbf{np} = [0; -1]$. The AFP equation is given by

$$\mathbf{vp}_0 \times (\mathbf{u} - \mathbf{q}_2) = 0. \quad (13)$$

We have $\mathbf{vp}_1 = \frac{1}{\mu_1}\mathbf{vp}_0 + b_1\mathbf{np}$. From (3),

$$b_1 = (d_0 - \sqrt{d_0^2 - (1-\mu_1^2)(x^2+d_0^2)})/\mu_1. \quad (14)$$

The refraction point $\mathbf{q}_2 = [x, d_0] - d_1\mathbf{vp}_1/c_1$. Substituting all these terms in (13), we get

$$(d_0^2\mu_1^2 + \mu_1^2x^2 - x^2)(d_0u^x + d_1x - u^yx)^2 = (d_0d_1x)^2. \quad (15)$$

Similar to Case 1, we obtain a 4^{th} degree equation in x. This gives four solutions for x and the correct solution is found by checking Snell's law for each solution. After obtaining x, the camera ray is obtained as $x\mathbf{z}_2 + d_0\mathbf{z}_1$ using which the image projection $\widehat{p}$ can be computed via internal camera calibration matrix. In Section 4, we showed that d_0 cannot be estimated for Case 2. The reader might wonder how AFP equation can then be solved. The key idea is that the camera ray can be computed by using *any* value of d_0. For example, let $d_0^{'} = \lambda d_0$ for $\lambda > 0$. Then the solution $x^{'} = \lambda x$. Thus, the camera ray remains the same. For Case 3, given calibration parameters d_0, d_1, μ_1, μ_2 and the known 3D point, a 12^{th} degree AFP equation can be derived similarly[1].

Iterative Refinement: Let $\mathbf{x}_c$ denote all calibration parameters. Given an initial estimate of $\mathbf{x}_c$ and pose, let $\widehat{p}(i) = \texttt{AFP}(\mathbf{x}_c, R\mathbf{P}(i) + \mathbf{t})$ be the image projection of $\mathbf{P}(i)$ computed by solving the AFP equation. The reprojection error is defined as the root mean square (RMS) error $J = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(p(i) - \widehat{p}(i))^2}$. We use `lsqnonlin` in Matlab to refine $\mathbf{x}_c$ and pose $[R, \mathbf{t}]$ by minimizing J.

5.1. Simulations

Now we show simulations for complete calibration and pose estimation using the same settings as in Section 3.2. The 8pt algorithm is used since it works better. In the RANSAC framework, after estimating the axis, the best set of 8 points are used to compute α and d_i's in a least square fashion as described in Section 4. Since there are 4 solutions for R from E matrix, we get 4 solutions for α and d_i's. The correct solution is found by enforcing $\alpha > \sum d_i$ and $d_i > 0 \; \forall \; i$. The obtained initial solution is refined by minimizing the reprojection error using the AFP. We also compute the pose obtained using a central (perspective) approximation from the given 2D-3D correspondences (referred by

[1] See supplementary materials.

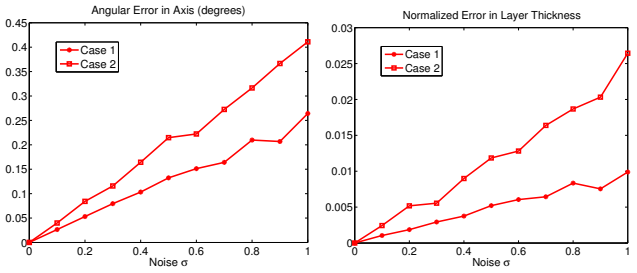


Figure 4. Error in axis and layer thickness for Case 1 and Case 2 using 8pt algorithm after non-linear refinement.

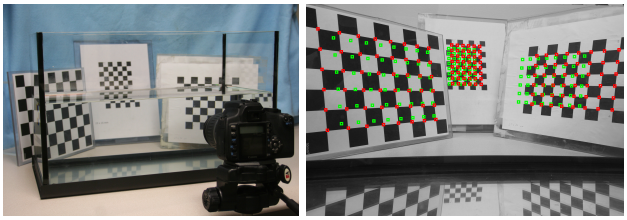


Figure 5. (Left) Setup. (Right) Photo captured by looking through a water tank. Projected 3D points are overlayed by applying pose estimated using CA (green) and our algorithm (red).

CA). Figure 3 and 4 show error plots for pose and calibration parameters and the final reprojection error for different noise levels, averaged over 100 trials. These plots show that correct calibration and pose parameters can be obtained using our algorithm. Notice the large translation and reprojection error, and smaller estimated $\mathbf{t}_z$ when using a central approximation (CA) in Figure 3. This is expected, because when looking through a denser refractive medium, objects appear *closer* to the camera. Note that the error due to noise in CA is insignificant compared to the error due to incorrect modeling.

6. Real Experiments

We show real results using a water tank of dimensions $508 \times 260 \times 300$ mm^3. We use a Canon Rebel XT camera having resolution of 3456×2304 pixels with a $18-55$ mm zoom lens. The camera was internally calibrated offline. Figure 5 shows a photo of a scene consisting of three checkerboards, captured by looking through the water tank (facing 260 mm side of tank). In order to obtain ground truth, we took another photo in air, using which the extrinsic of checkerboards were computed. The resulting 3D points in the coordinate system of left checkerboard are shown in Figure 6. We detect corners in the captured photo and run our algorithm (Case 2) to estimate the calibration and pose parameters. Figure 6 shows the estimated rotation and translation parameters, along with the final reprojection error J. The estimated thickness of the tank using our algorithm was 255.69 mm, resulting in a relative error of 1.66%. Notice the large error in $\mathbf{t}_z$ and large J in central approximation, also evident from projected points in Figure 5. Interestingly, the central approximation can recover the rotation well enough.

6.1. Planar Calibration Grid

We now show that calibration can also be done using a *single* planar grid, which is useful in practice. We describe an 8pt algorithm as follows. Without loss of generality, assume the plane is aligned with xy plane ($\mathbf{P}^z(i) = 0$). Substituting in the coplanarity equation (6), the columns 7, 8, 9 of B matrix reduce to zero. Let B' be the reduced 8×9 matrix, whose rank is 8. Thus, we can directly estimate the first two columns of the E matrix and $\mathbf{s}$ by SVD based solution using 8 correspondences. The last column of E is recovered using Demazure constraints [4] and $det(E) = 0$ constraint.

Let $E = \begin{bmatrix} e_1 & e_4 & x \\ e_2 & e_5 & y \\ e_3 & e_6 & z \end{bmatrix}$, where e_i's are estimated as above and x, y, z are unknown. Setting $det(E) = 0$ gives a linear equation using which x can be obtained as $x = ((e_1e_6 - e_3e_4)y + (e_2e_4 - e_1e_5)z)/(e_2e_6 - e_3e_5)$. The Demazure constraints provide three cubic equations and six quadratic equations in unknowns, from which any two quadratic equations can be chosen. Substituting x results in two quadratic equations in y and z, which can be solved to obtain a 4^{th} degree equation in z. In general, there are two real solutions which differ in sign. Thus, we obtain a pair of E matrices which differ in sign of their last column. Each pair of obtained rotation matrices also have the same property. The correct rotation matrix is chosen by checking for the determinant value of one[2].

Figure 7 shows simulation results on estimating calibration and pose parameters for Case 1 and Case 2 using a planar grid. Again, note that a central approximation completely fails. For real data shown in Figure 5, we estimated the calibration parameters using only the left checkerboard as shown in Figure 6. Thus, we see that calibration can also be performed using a single planar grid.

6.2. Unknown Refractive Indices

Due to lack of space, we only consider Case 1. We have three unknowns d_0, μ_1 and α. When $\mu_i's$ are unknown, ray directions cannot be pre-computed and FRC needs to be written in terms of camera rays as follows

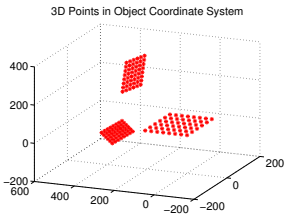
$$(a_1\mathbf{vp}_0 + b_1\mathbf{z}_1) \times (\mathbf{u} + \alpha\mathbf{z}_p + d_0\mathbf{vp}_0/c_0) = 0. \quad (16)$$

Let $\mathbf{vp}_0 = [v^x; v^y]$. After substituting a_1 and b_1 and removing the square root term in b_1, we get

$$(d_0v^x - v^yu^x)(\gamma + (v^y)^2 - 1) = (v^xv^y(\alpha - d_0 - u^y))^2, \quad (17)$$

where $\gamma = \mu_1^2$. Let $[EQ_i]_{i=1}^3$ be the 3 equations for 3 correspondences. Using EQ_1, γ can be obtained as a function of d_0 and α. Substituting γ in EQ_2 and EQ_3 makes

[2]The determinant of incorrect rotation matrices equals -1, corresponding to a reflection.



	N	$\theta_x, \theta_y, \theta_z$ (deg)	t_x, t_y, t_z (mm)	J (pixel)	d_1 (mm)
GT		$131.38, 1.22, 84.07$	$-237.58, -128.85, 455.80$	0	260
CA	144	$130.24, 1.42, 83.84$	$-217.71, -120.73, 372.14$	10.1	-
Ours†§	144	$131.38, 1.26, 84.12$	$-237.11, -128.16, 453.12$	0.33	255.69
Ours$^{\dagger\sharp}$	48	$131.40, 1.36, 84.03$	$-239.76, -129.26, 456.34$	0.26	272.81
Ours‡	144	$131.37, 1.26, 84.12$	$-236.46, -127.86, 449.70$	0.33	262.39

$\dagger$: Assuming $\mu_1 = 1.33$, $\S$ All Planes, $\sharp$ Left Plane Only, $\ddagger$ All Planes, Unknown μ_1

Figure 6. (Left) 3D points in the left checkerboard coordinate system. (Right) Estimates of pose and water-tank thickness d_1, and the final reprojection error J for real data shown in Figure 5 using central approximation (CA) and our algorithm. GT denotes ground-truth and N denotes the number of 2D-3D correspondences used.

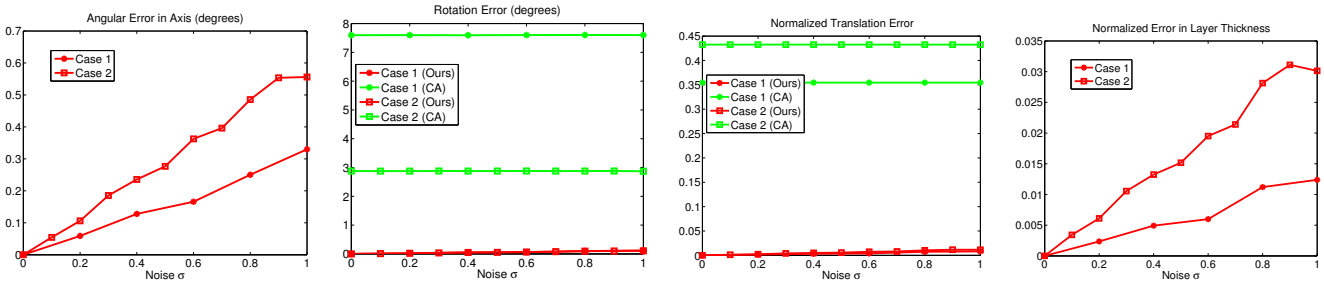


Figure 7. Errors in axis, rotation, translation and layer thickness using a planar calibration grid for different noise values, averaged over 100 trials. Rotation and translation errors using a central approximation (CA) are also shown.

them cubic in d_0 and quadratic in α. After eliminating α^2 between EQ_2 and EQ_3, α is obtained as a cubic function of d_0, which when substituted back into EQ_3 results in a 6^{th} degree equation in one unknown d_0. Solving it results in 6 solutions. The correct solution is found by enforcing $d_0 > 0$, $\alpha > d_0$ and $\mu_1 > 0$. Similarly, Case 2 also results in a 6^{th} degree equation[3]. However, Case 3 proved too difficult to obtain an analytical equation. Thus, multi-layer systems require good initial guess when μ_i's are unknown. Figure 6 shows the pose and calibration estimates for real data (Figure 5) assuming unknown μ_1 for water, which was recovered as 1.296 (relative error 2.55%).

7. Analysis

Field-of-View (FOV): Small FOV results in ambiguity between α and layer thicknesses as analyzed below using small angle approximation $\sin\theta \sim \tan\theta \sim \theta$. For small angles, $\theta_i \sim \theta_0/\mu_i$, where θ_0 is the angle between the camera ray and the axis[4]. Rewriting the FRC (9) on the POR using angles,

$$\tan\theta_n = (u^x - \sum d_i \tan\theta_i)/(u^y + \alpha - \sum d_i). \quad (18)$$

$$\theta_0 \overbrace{(\alpha + \mu_n \sum \frac{d_i}{\mu_i} - \sum d_i)}^{\chi} = -u^y\theta_0 + u^x\mu_n. \quad (19)$$

Thus, even when μ_i's are known, the only quantity that can be estimated is χ, which is a combination of α and d_i's. This also implies that even when the depth of the calibration object (and hence α) is known as in [25], individual layer thickness cannot be obtained for small FOV. In Figure 5, if we perform calibration using the center checkerboard only, the rotation error is within 0.5° whereas the tank thickness d_1 is estimated to be 550.38 mm (error of 290.38 mm), along with 70.45 mm error in α. However, the error in corresponding quantity $\chi = \alpha + d_1(\frac{1}{\mu_1} - 1)$ from the true value is only 1.59 mm. A central approximation in this case also gives a low reprojection error with similar translation error.

Multi-Layer Refractions: Non-central cameras can be well modeled using central approximation when the locus of viewpoints is small (*e.g.* catadioptric camera with small mirror compared to scene depths). For multi-layer refractions, a natural question then arises whether they can be modeled by simpler single/two-layer models. This also becomes important when multi-layer modeling becomes challenging due to similar or unknown refractive indices. We analyze if Case 3 (air-glass-water) can be approximated by Case 1 (air-medium). We perform simulation as in Section 3.2 using $\mu_1 = 1.5$, $\mu_2 = 1.33$, $d_0 = 300$ units and $d_1 = 450$ units. We add different amount of noise in 2D features and perform 100 trials for each. For each trial, we apply central approximation, single layer (SL) approximation using $\mu_1^{SL} = 0.5(\mu_1 + \mu_2) = 1.415$, and the correct two-layer model to estimate the pose and calibration parameters. Figure 8 shows the average error plots. Notice that with no noise, both SL and CA approximations give non-zero errors while the correct model gives zero error. However, as noise is increased, SL approximation gives similar pose, axis and reprojection errors compared to the true model. As expected, a central approximation is significantly worse than SL approximation. In general, with large noise

[3] Supplementary materials provide details for Case 1 and Case 2.

[4] Assuming small angle between camera's optical axis and normal $\mathbf{n}$.

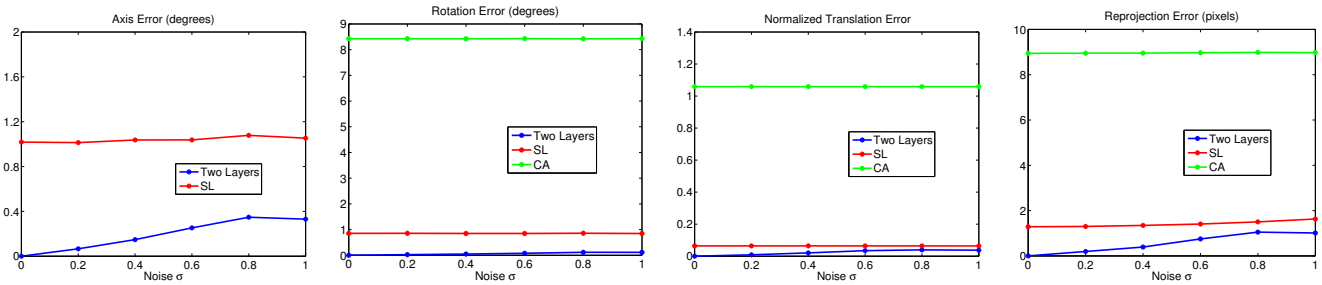


Figure 8. As noise increases, a two layer refraction system (Case 3) can be well-approximated using a single layer system (Case 1), but not using a central approximation.

we expect multi-layer refractions to be modeled well by either Case 1 or Case 2 depending on $\mu_n \neq \mu_0$ or $\mu_n = \mu_0$ respectively.

8. Conclusions

We have analyzed the geometry of a perspective camera imaging through multiple flat refractive layers. We developed a theory for calibration and derived forward projection equations, which can be directly used in applications such as 3D reconstruction [1]. We presented a comprehensive analysis under unknown layer distances and orientation, and known/unknown refractive indices. Since calibration can be done using a single planar grid, the proposed algorithms are useful in practical scenarios such as underwater imaging. We showed that multi-layer systems may be well-approximated by simpler single layer systems. Multiple planar grids can be used to increase the calibration accuracy similar to calibration of perspective cameras. Our proposed 8-point algorithm for axis computation can be used for other axial setups such as catadioptric cameras, as well as to compute the distortion center for fish-eye cameras [7]. Developing a minimal solution for calibrating flat refractive geometry remains an interesting future work.

References

[1] Y.-J. Chang and T. Chen. Multi-view 3D reconstruction for scenes under the refractive plane with known vertical direction. In *ICCV*, 2011. 2, 8

[2] V. Chari and P. Sturm. Multiple-view geometry of the refractive plane. In *BMVC*, Sep 2009. 2

[3] Z. Chen, K.-Y. K. Wong, Y. Matsushita, X. Zhu, and M. Liu. Self-calibrating depth from refraction. In *ICCV*, 2011. 2

[4] M. Demazure. Sur deux problemes de reconstruction. Technical Report 882, INRIA, 1988. 6

[5] G. Glaeser and H.-P. Schröcker. Reflections on refractions. *J. Geometry and Graphics*, 4(1):1–18, 2000. 2, 5

[6] A. S. Glassner. *An Introduction to Ray Tracing*. Morgan Kaufmann, 1989. 3

[7] R. I. Hartley and S. B. Kang. Parameter-free radial distortion correction with center of distortion estimation. *PAMI*, 29(8):1309–1321, Aug. 2007. 3, 8

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 3

[9] J. Höhle. Reconstruction of the underwater object. *Photogrammetric Engineering*, pages 948–954, 1971. 1

[10] R. Kotowski. Phototriangulation in multi-media photogrammetry. *Int'l Archives of Photogrammetry and Remote Sensing*, XXVII, 1988. 2

[11] H. Li, R. Hartley, and J. Kim. A linear approach to motion estimation using generalized camera models. In *CVPR*, 2008. 2

[12] R. Li, H. Li, W. Zou, R. G. Smith, and T. A. Curran. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE J. Oceanic Eng.*, 2:364–375, 1997. 2

[13] H.-G. Maas. New developments in multimedia photogrammetry. *Optical 3D Measurement Techniques*, III, 1995. 2

[14] N. Morris and K. Kutulakos. Dynamic refraction stereo. In *ICCV*, volume 2, pages 1573–1580, 2005. 2

[15] D. Nistér. An efficient solution to the five-point relative pose problem. *PAMI*, 26(6):756–770, June 2004. 1, 3

[16] R. Pless. Using many cameras as one. In *CVPR*, pages 587–594, 2003. 2

[17] S. Ramalingam, P. Sturm, and S. K. Lodha. Theory and calibration algorithms for axial cameras. In *ACCV*, 2006. 2

[18] K. Rinner. Problems of two-medium photogrammetry. *Photogrammetric Engineering*, 35(3):275–282, 1969. 1

[19] M. Shimizu and M. Okutomi. Calibration and rectification for reflection stereo. In *CVPR*, pages 1–8, June 2008. 2

[20] M. Shortis, E. Harvey, and J. Seager. A review of the status and trends in underwater videometric measurement. In *SPIE Conf. 6491, Videometrics IX*, Jan. 2007. 1

[21] E. Steger and K. Kutulakos. A theory of refractive and specular 3D shape by light-path triangulation. *IJCV*, 76(1):13–29, 2008. 2, 5

[22] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *ECCV*, 2004. 2

[23] J.-P. Tardif, P. Sturm, M. Trudeau, and S. Roy. Calibration of cameras with radially symmetric distortion. *PAMI*, 31(9):1552–1566, 2009. 1, 2

[24] S. Thirthala and M. Pollefeys. Multi-view geometry of 1D radial cameras and its application to omnidirectional camera calibration. In *ICCV*, volume 2, pages 1539–1546, 2005. 2

[25] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. In *CVPR*, 2008. 1, 2, 7