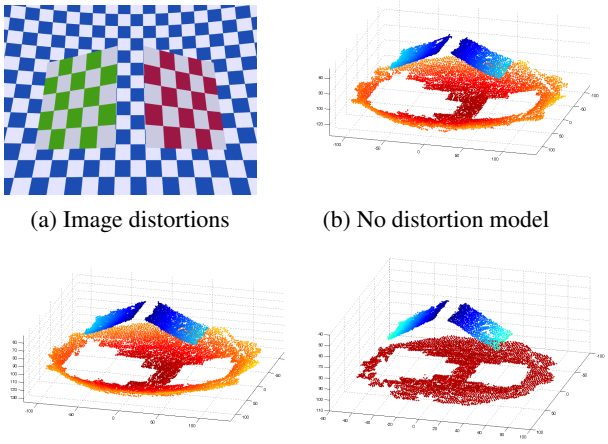
Multi-View 3D Reconstruction for Scenes under the Refractive Plane with Known Vertical Direction

Yao-Jen Chang, Tsuhan Chen
Cornell University
{ychang, tsuhan}@cornell.edu

Abstract

Images taken from scenes under water suffer distortion due to refraction. While refraction causes magnification with mild distortion on the observed images, severe distortions in geometry reconstruction would be resulted if the refractive distortion is not properly handled. Different from the radial distortion model, the refractive distortion depends on the scene depth seen from each light ray as well as the camera pose relative to the refractive surface. Therefore, it's crucial to obtain a good estimate of scene depth, camera pose and optical center to alleviate the impact of refractive distortion. In this work, we formulate the forward and back projections of light rays involving a refractive plane for the perspective camera model by explicitly modeling refractive distortion as a function of depth. Furthermore, for cameras with an inertial measurement unit (IMU), we show that a linear solution to the relative pose and a closed-form solution to the absolute pose can be derived with known camera vertical directions. We incorporate our formulations with the general structure from motion framework followed by the patch-based multiview stereo algorithm to obtain a 3D reconstruction of the scene. We show through experiments that the explicit modeling of depth-dependent refractive distortion physically leads to more accurate scene reconstructions.



(a) Image distortions (b) No distortion model

(c) Radial distortion model (d) Refractive distortion model

Figure 1. 3D reconstruction for a synthetic scene with three planar surfaces under water: (a) An input image of the scene where the surfaces are textured with checkerboard patterns to demonstrate the refractive distortion. (b) 3D scene reconstruction without distortion modeling, resulting in noticeable curved reconstruction for the bottom plane and slightly curved reconstruction on the other two planes. (c) The radial distortion model provides minor help in recovering the geometry distortion for 3D reconstruction. (d) 3D scene reconstruction obtained by incorporating the refractive distortion, where all surfaces are correctly reconstructed as planes. The color in the scene reconstruction encodes the height of a point from the lowest (red) to the highest (blue). (Best viewed in color.)

1. Introduction

Refraction of light is a commonly observed phenomenon where the light changes its direction due to an alternation of the propagation speed in different mediums. This results in noticeable distortion when seeing things through a transparent medium like water, in which people observe illusions of an underwater scene to become closer to the surface and get magnified. To reconstruct the underwater scene, however, is not trivial even for scenes under a flat water surface, where only mild image distortion is generated. This is because the refractive distortion depends on not only the distance from the optical axis, as commonly used for modeling lens radial distortion, but also the scene depth of each light ray as well as the camera position and orientation relative to the refractive plane. As shown in Figure 1 with a synthetic example, if refractive distortion is not taken into consideration or simply modeled as radial distortion, incorrect 3D scene structure will be resulted. Therefore, explicit modeling of the refractive distortion as a function of scene depth and camera parameters is crucial for scene reconstruction under the refractive plane.

In this work, we focused on the scene reconstruction un-

der a flat refractive geometry as it is one of the most representative settings for scenes involving refraction. Typical cases include (1) flat surfaces separating the viewer and the water body in aquaria, and (2) cameras enclosed in an air chamber immersed in the water for underwater photography. As addressed in [26], the analysis of this setting not only contributes to the vision challenges it poses, but also extends computer vision to a variety of applications in oceanic engineering, psychology, biology, and archaeology.

For most of the underwater photography work, the camera setup is fixed and therefore camera positions and orientations relative to the refractive plane can be measured with known calibration objects [17]. Here we deal with a more general multi-camera setting and propose to utilize the scene itself to obtain the estimate of camera parameters and scene depth. This naturally leads to a new formulation of structure and motion estimation framework involving the refractive distortion. Our primary contributions are:

- With a thorough analysis of the basic components of structure and motion estimation, an analytical formulation is presented for forward projection and back projection by incorporating the refractive distortion.
- By explicitly incorporating the depth-dependent ratio term into projection modeling, a simple yet exact two-step formulation can be derived by separating the refractive and perspective projections, which also contributes to a much simpler form for deriving Jacobian for non-linear optimization within the bundle adjustment framework.
- For cameras with an inertial measurement unit (IMU), we show that a linear solution to the relative pose and a closed-form solution to the absolute pose can be derived with known camera vertical directions.

2. Related Work

Photogrammetry: The analysis of refractive geometry can be traced back to early work in photogrammetry started by Rinner [22], where the highly nonlinear refractive distortion was approximated by a number of differential linear transformations for different zones of scene depth. In 1970s, the usage of stereo cameras for underwater photography was addressed by Höhle [13], where a numerical method was described to find the unknowns of the refractive geometry. The successive efforts along the line of photogrammetry were either (1) trying to model the light path through multiple mediums with ray tracing [15], but requiring prior knowledge of the shape and position of the refractive surfaces or two-phase calibration with a known calibration frame [17], (2) seeking for efficient approximation to reduce the computation load as simple table lookup [19], or (3) simply allowing the refractive effects to be absorbed by the conventional camera calibration with a radial distortion model [23]. While these methods were designated for underwater photography with fixed refractive geometry, the more general multiview setting including multiple cameras or a camera with varying pose relative to the refractive surfaces has been less explored.

Image restoration and surface reconstruction: Image restoration under refractive distortion has been investigated since the early work of Hurase [21] for water surface reconstruction. A static camera with the orthographic projection model was assumed to observe an unknown planar pattern under a disturbing water surface. Optical flow was utilized to track point trajectory, and the center of a point trajectory is considered to stay on a flat water surface, thereby recovering the underwater pattern. Many works in image restoration followed similar experimental setting [8, 25], in which orthographic cameras and underwater planar patterns were assumed. With a known underwater planar pattern, light path triangulation could be performed to estimate the dynamic surface normals by using two cameras by Morris and Kutulakos [20], or higher-order curvature characteristics by using general linear cameras (GLC) approximation with a orthographic camera proposed by [6]. While orthographic camera is a useful model for surface reconstruction, this model may not be suitable in a setting where cameras are very close to the refractive surface as in many underwater applications, or 3D scene reconstruction is of interest.

3D reconstruction and depth estimation: Ben-Ezra and Nayar [3] proposed a model-based approach to recover the shapes and poses of transparent objects with known motion. By tracking points refracted through the transparent object, object parameters could be estimated within a set of object categories. For general scene reconstruction, Ding and Yu introduced epsilon stereo matching [5] for inferring 2D disparities by fusing two GLCs in an energy minimization framework. A volumetric reconstruction approach was further proposed [7] to derive more accurate 3D scene reconstruction. The reflective surface and the camera within the multiperspective camera system had to be calibrated in order to decompose the reflected surface into piecewise GLCs. Similar ideas can be applied for 3D reconstruction with the refractive plane when camera poses relative to the refractive plane have been estimated.

Multiview refractive geometry: The refractive geometry involving refraction with a perspective camera has not attracted much attentions in computer vision community until recently. In [26], Treibitz *et al*. categorized the common flat-interface for cameras in water-based applications as non single-viewpoint cameras. This indicates that the refractive distortion in general cannot be modeled as radial distortion when the camera pupil doesn't lie on the surface. Chari and Sturm [4] explored the refractive geome-

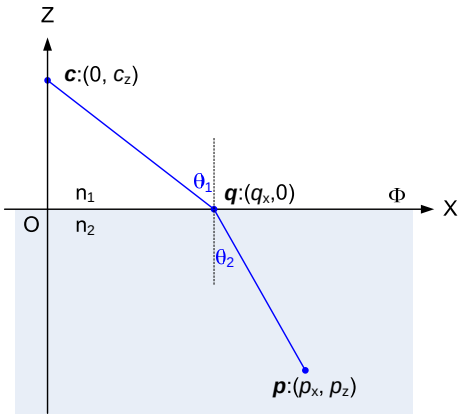


Figure 2. The refraction of a light in the 2D case.

try in a multi-camera setting, showing that theoretically the existence of geometric entities like the projection matrix, fundamental matrix and homography matrix. While it is of theoretical interest, we try to incorporate refractive distortion seamlessly within the conventional structure from motion framework, which leads to a simpler form ready for use. In [1, 2], Agrawal *et al*. analyzed forward projection for non-central catadioptric system comprising a perspective camera with multiple rotationally symmetric conic or quadric reflectors on a plane. While they utilized the contours of reflectors for initial pose estimation, and relied on bundle adjustment for non-linear optimization, we focus on the setting of multiple arbitrarily placed perspective cameras with a refractive plane, and use the proposed relative and absolute pose algorithms to estimate the initial pose. In the next section, we will present our formulation of fundamental elements that constitute the building blocks for the 3D reconstruction for scenes under the refractive plane.

3. The Algorithm

In this section, we start with describing the fundamental refraction law in the simplified 2D case, where a 4-th order equation is formulated for deriving the intersection point of light path and the refractive plane. We show that it naturally extends to the general 3D case by introducing the refrax ratio, which simplifies the derivation of forward projection by dividing the refractive projection into two parts. For 3D scene reconstruction, we take a similar approach as Snavely *et al*. [24] by firstly performing relative pose estimation for two images with a sufficient baseline. Next, we add one image at a time by using the absolute pose estimation. Initial estimate of 3D points are estimated by triangulation, followed by the bundle adjustment for joint optimization of structure and motion with scenes under a refractive plane. With the accurate estimate of camera parameters, the patch-based multiview stereo algorithm is incorporated for dense 3D scene reconstruction.

3.1. The refraction law

The refraction is governed by the Snell's law to relate the light paths of incident light and refracted light with respect to the surface normal of the refractive plane:

$$\sin\theta_1 = \delta \sin\theta_2, \tag{1}$$

where $\delta = n_2/n_1$ is the fraction ratio between the refractive indices of two mediums, θ_1 and θ_2 are the angles of the incident light and refractive light with respect to the surface normal as shown in Figure 2. The trigonometric relations can be represented with the geometric locations of optical center $\mathbf{c}$, and the point $\mathbf{p}$ under the refractive surface Φ. In the simplified 2D case, without loss of generality, let's assume the x-axis is aligned with the refractive surface, z-axis is aligned with the camera optical center, therefore $\mathbf{c} = (0, c_z)^\top$. With the point $\mathbf{p} = (p_x, p_z)^\top$, a fourth order quartic equation can be derived for the location of the intersection point $\mathbf{q} = (q_x, 0)^\top$ of the light path from $\mathbf{p}$ to $\mathbf{c}$ and the surface Φ [11]:

$$\begin{aligned} f(q_x) =& N q_x^4 - 2N p_x q_x^3 + \left(N p_x^2 - \frac{p_z^2}{\delta^2} + c_z^2 \right) q_x^2 \\ & - 2c_z^2 p_x q_x - c_z^2 p_x^2 = 0, \end{aligned} \tag{2}$$

with $N = 1 - \frac{1}{\delta^2}$. The point $\mathbf{q}$ is also called the refrax of $\mathbf{p}$ on the surface Φ.

3.2. The refrax ratio and forward projection

Before extending the 2D case into the general 3D case, we introduce the *refrax ratio* λ, which is the ratio between the projected distance from $\mathbf{c}$ to $\mathbf{q}$ onto the surface Φ, and the projected distance from $\mathbf{c}$ to $\mathbf{p}$ onto the surface Φ. That is,

$$\lambda = \frac{q_x - c_x}{p_x - c_x}. \tag{3}$$

With which, Eq. 2 can be rewritten as

$$\begin{aligned} f(\lambda) =& N a^4 \lambda^4 - 2N a^4 \lambda^3 + \left(N a^2 - \frac{p_z^2}{\delta^2} + c_z^2 \right) a^2 \lambda^2 \\ & - 2a^2 c_z^2 \lambda - a^2 c_z^2 = 0, \end{aligned} \tag{4}$$

where a is the projected distance between $\mathbf{c}$ and $\mathbf{p}$ onto the refractive surface Φ. The merit of using the refrax ratio is that the Eq. 4 also holds in the general 3D case. Let $\mathbf{c} = (c_x, c_y, c_z)^\top$ and $\mathbf{p} = (p_x, p_y, p_z)^\top$, the distance between the projections of $\mathbf{c}$ and $\mathbf{p}$ on the surface Φ becomes $a = \sqrt{(p_x - c_x)^2 + (p_y - c_y)^2}$. By solving λ in Eq. 4 given $\mathbf{c}$ and $\mathbf{p}$, the refrax $\mathbf{q}$ of $\mathbf{p}$ on the surface Φ would be

$$\mathbf{q} = (c_x + \lambda(p_x - c_x), c_y + \lambda(p_y - c_y), 0)^\top. \tag{5}$$

Since the light path from $\mathbf{q}$ to $\mathbf{c}$ is within a single medium (i.e., the air), conventional perspective projection can be

readily applied. Hence the forward projection of $\mathbf{p}$ onto the image plane of $\mathbf{c}$ via the surface Φ is simply achieved by projecting $\mathbf{q}$ to the image plane of $\mathbf{c}$ with perspective projection. The introduction of the refrax ratio λ simplifies not only the formulation of forward projection, but also the derivation of Jacobians in the later stage when we jointly optimize the camera poses, optical centers, and 3D points with the bundle adjustment framework as described in Section 3.5.

3.3. Back projection

As shown in the previous subsection, the relation between a 3D point under the refractive plane and its projection to a camera is highly nonlinear. The forward projection involves solving a quartic equation followed by conventional perspective projection. In contrast, back projection is relative simple by casting a ray from the camera optical center through the point projection on the image plane to reach the refractive surface Φ, and then following the Snell's law to derive the corresponding refractive light path under the surface. Denote the camera intrinsic matrix as $\mathbf{K}$ and extrinsic matrix as $\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}$, where $\mathbf{R} = \begin{bmatrix}\mathbf{r_1} & \mathbf{r_2} & \mathbf{r_3}\end{bmatrix}$ is the rotation matrix and $\mathbf{t}$ is the translation vector related with the optical center by $\mathbf{t} = -\mathbf{R}\mathbf{c}$. Given a projection point on the image plane denoted as $\mathbf{u} = (u, v, 1)^\top$, the refrax point $\mathbf{q} = (q_x, q_y, q_z)^\top$ can be derived by back projection:

$$\mathbf{q} = \mathbf{c} + w\mathbf{R}^\top\mathbf{K}^{-1}\mathbf{u}, \tag{6}$$

where the weighting factor w can be obtained using the fact that $\mathbf{q}$ lies on the surface Φ coinciding the xy-plane with $z = 0$. To simplify the notations, let's denote $\mathbf{s} = (s_x, s_y, s_z)^\top = \mathbf{R}^\top\mathbf{K}^{-1}\mathbf{u}$. Thus, $q_x = c_x - c_z s_x/s_z$, and $q_y = c_y - c_z s_y/s_z$. Let b be the projected distance between $\mathbf{q}$ and $\mathbf{c}$ onto the surface Φ, we have

$$b = \sqrt{(q_x - c_x)^2 + (q_y - c_y)^2} = |c_z|\sqrt{(s_x^2 + s_y^2)/s_z^2}. \tag{7}$$

A point $\mathbf{p} = (p_x, p_y, p_z)^\top$ on the refractive light can be written in terms of $\mathbf{q}$ and $\mathbf{c}$:

$$\begin{bmatrix}p_x \\ p_y\end{bmatrix} = \begin{bmatrix}q_x \\ q_y\end{bmatrix} + \frac{p_z}{\sqrt{(\delta^2 - 1)b^2 + \delta^2 c_z^2}}\begin{bmatrix}q_x - c_x \\ q_y - c_y\end{bmatrix} \tag{8}$$

$$= \begin{bmatrix}c_x \\ c_y\end{bmatrix} - \frac{1}{s_z}\begin{bmatrix}s_x \\ s_y\end{bmatrix}\left(c_z - \frac{p_z}{g(\mathbf{s})}\right), \tag{9}$$

where $g(\mathbf{s}) = \sqrt{(\delta^2 - 1)(s_x^2 + s_y^2)/s_z^2 + \delta^2}$ is a function dependent on the rotation matrix $\mathbf{R}$ and 2D projection $\mathbf{u}$, but independent of $\mathbf{c}$ and $\mathbf{p}$.

The formulation of back projection with Eq. 9 leads to a linear relation between the camera center and the 3D point via its 2D projections given the camera pose. This is the case for triangulation, where the 3D location of a point $\mathbf{p}$ can be estimated by its 2D projections on two or more cameras with known intrinsic and extrinsic parameters. For the cases where only partial pose information is available such as cameras with an inertial measurement unit (IMU), the back projection can be reformulated to form a linear solution for relative pose estimation and a closed-form solution for absolute pose estimation with two points, which will be addressed in the next subsection.

3.4. Pose estimation with known vertical direction

Thanks to the recent advances of MEMS technology, the deployment of MEMS-IMU in consumer electronic devices such as digital cameras and smart phones is getting more and more pervasive. IMUs such as accelerometers and digital compasses can be used to measure the orientation of the device. Although the accuracy of the heading measured by the digital compass is not good enough due to magnetic filed disturbances, the roll and pitch angles can be measured accurately even with low-cost IMUs [16]. Therefore, attempts have been made to take the advantage of two known angles (i.e., the vertical direction) to simplify the relative pose estimation [9, 14] and absolute pose estimation [16] for settings involving only one medium.

For settings involving a refractive plane like water, conventional epipolar geometry does not exist. Chari and Sturm [4] derived a fundamental matrix of dimensions 12×12 that relates the lifted coordinates in one image to a quartic curve in the other image. With the known vertical direction, we will show that the estimation can be largely simplified as the incident angle of each light ray from the optical center passing through each image point becomes available for calibrated cameras.

3.4.1 Relative pose estimation

For relative pose estimation with known vertical direction, the task is to estimate relative position and relative heading of the second camera with respect to the first camera given a set of 2D point correspondences. By decomposing the rotation matrix $\mathbf{R}$ by the rotation $\mathbf{R}_z$ around Z-axis and the rotation $\mathbf{R}_{ver}$ formed by the vertical direction, the refrax point $\mathbf{q}$ in Eq. 6 can be rewritten as:

$$\mathbf{q} = \mathbf{c} + w\mathbf{R}_z^\top\mathbf{R}_{ver}^\top\mathbf{K}^{-1}\mathbf{u} \tag{10}$$

$$= \mathbf{c} - \frac{c_z}{v_z}\mathbf{R}_z^\top\mathbf{v}, \tag{11}$$

where $\mathbf{v} = (v_x, v_y, v_z)^\top = \mathbf{R}_{ver}^\top\mathbf{K}^{-1}\mathbf{u}$ is the transformed ray direction of the image point $\mathbf{u}$ with zero pitch and zero roll with respect to the optical center, and

$$\mathbf{R}_z = \begin{bmatrix}\mathcal{C}_\phi & -\mathcal{S}_\phi & 0 \\ \mathcal{S}_\phi & \mathcal{C}_\phi & 0 \\ 0 & 0 & 1\end{bmatrix}, \tag{12}$$

where we denote $\mathcal{C}_\phi = \cos\phi$ and $\mathcal{S}_\phi = \sin\phi$. With above parameterizations, the relation in Eq. 9 between the 3D point $\mathbf{p}$ and optical center $\mathbf{c}$ can be described by

$$\begin{bmatrix} p_x \\ p_y \end{bmatrix} = \begin{bmatrix} q_x \\ q_y \end{bmatrix} + \frac{p_z}{v_z g(\mathbf{v})} \begin{bmatrix} v_x & v_y \\ v_y & -v_x \end{bmatrix} \begin{bmatrix} \mathcal{C}_\phi \\ \mathcal{S}_\phi \end{bmatrix} \tag{13}$$

$$= \begin{bmatrix} c_x \\ c_y \end{bmatrix} - \frac{1}{v_z}\left(c_z - \frac{p_z}{g(\mathbf{v})}\right) \begin{bmatrix} v_x & v_y \\ v_y & -v_x \end{bmatrix} \begin{bmatrix} \mathcal{C}_\phi \\ \mathcal{S}_\phi \end{bmatrix}, \tag{14}$$

where $g(\mathbf{v}) = \sqrt{(\delta^2-1)(v_x^2+v_y^2)/v_z^2 + \delta^2}$.

Without loss of generality, let's assume the optical center of the first camera is located at $\mathbf{c}_1 = (0, 0, c_{1z})^\top$ with zero heading ($\phi_1 = 0°$). We can express p_x and p_y with p_z:

$$\begin{bmatrix} p_x \\ p_y \end{bmatrix} = \frac{1}{v_{1z}}\left(\frac{p_z}{g(\mathbf{v}_1)} - c_{1z}\right) \begin{bmatrix} v_{1x} \\ v_{1y} \end{bmatrix}, \tag{15}$$

where $\mathbf{v}_1$ is the transformed ray direction of $\mathbf{u}_1$. By combining the above two equations with the transformed light ray $\mathbf{v}_2$ associated with the corresponding point $\mathbf{u}_2$ on the second image, we can eliminate p_z to get a linear equation $\mathbf{e}^\top \mathbf{x} = 0$, where the vector $\mathbf{e}$ is composed of entries with known variables $\mathbf{v}_1, \mathbf{v}_2, g(\mathbf{v_1}), g(\mathbf{v_2}), h$, and $\mathbf{x}$ is composed of a set of unknown variables $[\mathcal{C}_{\phi_2}, \mathcal{S}_{\phi_2}, c_{2z}\mathcal{C}_{\phi_2}, c_{2z}\mathcal{S}_{\phi_2}, c_{2x}\mathcal{C}_{\phi_2} + c_{2y}\mathcal{S}_{\phi_2}, c_{2x}\mathcal{S}_{\phi_2} - c_{2y}\mathcal{C}_{\phi_2}, c_{2x}, c_{2y}]$ with the constraint:

$$x_1^2 + x_2^2 = \mathcal{C}_{\phi_2}^2 + \mathcal{S}_{\phi_2}^2 = 1. \tag{16}$$

As there are eight unknown variables with one constraint, and each pair of point correspondences provide one linear equation, an equation system can be formed by stacking the vectors $\mathbf{e}$ with seven pairs of point correspondences:

$$\mathbf{E}\mathbf{x} = \mathbf{0}. \tag{17}$$

A solution can be obtained by scaling the eigenvector corresponding to the smallest generalized eigenvalue of the symmetric matrix $\mathbf{E}^\top\mathbf{E}$ and the diagonal matrix $\mathbf{B} = \text{diag}([1, 1, 0, 0, 0, 0, 0, 0])$.

Note that there are two special cases that would cause the rank deficiency to Eq. 17: (1) All 3D points are on the plane passing through two camera centers and perpendicular to the refractive plane; (2) The two camera centers are on the same height, and all 3D points are on the bisecting plane of the line connecting two camera centers. These cases can hardly happen in reality, and they can be resolved by either selecting other point correspondences that don't result in rank deficiency within a RANSAC framework, or using other cameras as the initial pair for relative pose estimation.

3.4.2 Absolute pose estimation

In [16], Kukelova *et al*. presented a closed form solution for the absolute pose estimation with known vertical direction. A minimal case of 2 points is derived by using the following equality that relates a 3D point $\mathbf{p}$ and its 2D projection $\mathbf{u}$:

$$[\mathbf{u}]_\times \mathbf{K}[\mathbf{R}_{ver}\mathbf{R}_z|\mathbf{t}]\mathbf{p} = 0. \tag{18}$$

In the setting with the refractive plane, the above equality only holds for the refrax point $\mathbf{q}$. Therefore, we can use Eq. 13 to relate the 3D point $\mathbf{p}$ and its 2D projection $\mathbf{u}$:

$$[\mathbf{u}]_\times \mathbf{K}[\mathbf{R}_{ver}\mathbf{R}_z|\mathbf{t}] \begin{bmatrix} p_x - \frac{p_z(v_x\mathcal{C}_\phi + v_y\mathcal{S}_\phi)}{v_z g(\mathbf{v})} \\ p_y - \frac{p_z(v_y\mathcal{C}_\phi - v_x\mathcal{S}_\phi)}{v_z g(\mathbf{v})} \\ 0 \\ 1 \end{bmatrix} = 0. \tag{19}$$

With similar trick by modeling $\tau = \tan\frac{\phi}{2}$ as in [16], each point with Eq. 19 provides two linear equations with monomials $\tau^2, \tau, t_x, t_y, t_z, 1$. Therefore, with two points, four equations can be used to eliminate t_x, t_y, t_z, which lead to a second-order polynomial of τ with two solutions. By back-substituting the solutions to the original equations, we can obtain t_x, t_y, t_z and then test which solution provides smaller reprojection error.

While the minimal case provides efficient solutions in the RANSAC framework for removing the outliers, a linear formulation treating $\mathcal{C}_\phi$ and $\mathcal{S}_\phi$ as separate variables is required for absolute pose estimation with more than two points. This forms a constrained least squares problem with a quadratic inequality constraint (LSQI), which can be resolved by using the generalized singular value decomposition followed by root searching [12].

3.5. Bundle adjustment

While the solutions from the relative pose and absolute estimation are useful for estimating the locations of cameras and 3D points, the solution obtained from linear estimation doesn't necessary minimize the reprojection error. Furthermore, the camera pose obtained with IMUs may also contain small errors which can't be recovered in linear optimization. To handle these issues, a non-linear solution by bundle adjustment is investigated. Given m cameras and n 3D points with good initial guess, the goal of bundle adjustment is to adjust the camera parameters and 3D point positions to minimize the distance between measured projection point $\mathbf{u}_{ij}$ of the i-th 3D point $\mathbf{p}_i$ to the j-th camera $\mathbf{c}_j$, and the estimated projection $\hat{\mathbf{u}}_{ij}$ obtained with the estimated 3D points and cameras parameters, i.e.,

$$(\mathbf{C}^*, \mathbf{P}^*) = \underset{\mathbf{C},\mathbf{P}}{\text{argmin}} \sum_{i,j} \|\mathbf{u}_{ij} - \hat{\mathbf{u}}_{ij}\|_2^2, \tag{20}$$

where $\hat{\mathbf{u}}_{ij}$ is the image projection of the refrax point of the point $\mathbf{p}_i$ and camera center $\mathbf{c}_j$:

$$\begin{bmatrix} \hat{\mathbf{u}}_{ij} \\ 1 \end{bmatrix} \equiv \mathbf{K}\mathbf{R}(\hat{\mathbf{q}}_{ij} - \hat{\mathbf{c}}_j) = \mathbf{K}\mathbf{R} \begin{bmatrix} \hat{\lambda}_{ij}(\hat{p}_{ix} - \hat{c}_{jx}) \\ \hat{\lambda}_{ij}(\hat{p}_{iy} - \hat{c}_{jy}) \\ -\hat{c}_{jz} \end{bmatrix}, \tag{21}$$

where the refrax ratio $\hat{\lambda}_{ij}$ can be solved from the quartic equation in Eq. 4.

Unlike common practice for optimizing extrinsic camera parameters by its orientation and translation derived from the camera projection matrix, Eq. 21 leads us to direct optimization over the camera centers instead of translation vectors. In this way, the projection can be decomposed into the refractive projection between the mediums and the perspective projection above the surface. For refractive projection, the refrax ratio is independent of camera poses. For perspective projection, the refrax ratio can be treated as a constant. Therefore, the Jacobian of $\hat{\mathbf{u}}_{ij}$ with respect to camera parameters and 3D point positions can be easily formulated.

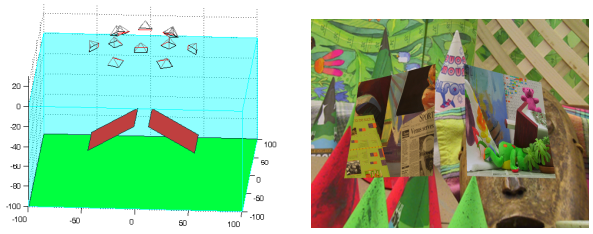
We incorporate the derived Jacobian into the sparse bundle adjustment framework [18, 27], which exploits the sparsity of the underlining block structure of normal equations for efficient optimization of the camera poses, optical centers, and 3D point positions. Accurate solutions can be obtained as validated in our synthetic experiments.

3.6. Patch-based scene reconstruction

In the bundle adjustment framework, sparse point correspondences by using SIFT features are sufficient for deriving accurate camera parameters. However, the 3D point cloud itself is rather sparse and insufficient to represent the scene geometry. To expand the reconstruction points, we adopt the patch-based multiview stereo (PMVS) algorithm by Furukawa and Ponce [10] to reconstruct a denser scene geometry based on the expansion of matched Harris and DoG interest points across multiple images. Forward projection and back projection as addressed above have to be incorporated in order to take refraction into account. The formulation of back projection also naturally extends to a linear solution of triangulation for estimating 3D point positions with known camera parameters and point correspondences. A better scene reconstruction can be obtained by incorporation of the refractive distortion with additional computation load contributed by solving the quartic equation for every forward projection. Note that the epipolar constraint is utilized in PMVS to collect candidates for point matching. As the conventional epipolar constraint doesn't hold for scenes with refractive distortion, this constraint needs to be relaxed so that matched features can be included in the collected candidates. This idea is similar to epsilon stereo matching proposed by Ding and Yu [5].

4. Experiment

To evaluate the proposed approach, experiments are conducted on both synthetic and real image sequences. In these experiments, we compare three different cases of distortion modeling: 1) no distortion model, 2) radial distortion model, and 3) our proposed refractive distortion model. For the first two cases, the structure from motion is performed with the Bundler [24] without considering the refraction. In the second case, the radial distortion of camera lens is optimized within the the non-linear optimization process for minimizing the reprojection error. In the third case, relative pose and absolute pose estimation are performed with the RANSAC framework to remove possible outliers. The nonlinear optimization by bundle adjustment is incorporated for accurate camera localization, pose estimation, and position estimation for the sparse point cloud.



(a) (b)

Figure 3. (a) The setup of the synthetic scene composed of 3 planar surfaces and 12 cameras. (b) One sample image from the synthetic image sequence.

Table 1. Camera Localization Error Comparison (cm)

	No distortion	Radial distortion	Refractive distortion
rms error	1.062	0.717	0.025

4.1. Synthetic data

We create a synthetic scene rendered by the POVRay ray tracing program for quantitative evaluation. The synthetic scene is composed of three planar surfaces with texture mapped from the images acquired from the Middlebury stereo dataset (Teddy, Cones, and Venus). The scene geometry is modeled as a square water tank with size $200 \times 200 \times 100$ (L×W×H cm), with one plane on the bottom, and the other two slanted planes tilted by $\pm 30^\circ$ as shown in Figure 3(a). The camera is placed above water with two different heights (30 and 40 cm, respectively), following a rounded path with 12 images taken in total. The camera field of view is set to 75°, with a pose facing the scene center. One sample image of the synthetic image sequence is shown in Figure 3(b), where the refractive distortion is easily observable.

First, we evaluate the performance of camera localization. As shown in Table 1, the proposed refractive distortion model greatly enhances the accuracy of camera localization. In the first two cases, since the camera localization is not aware of water, the camera locations could be up to a similarity transform to the ground truth. We align the camera locations with an estimate of similarity transform against the ground truth, and then calculate the displacement error. As expected, the error for the first case with no distortion

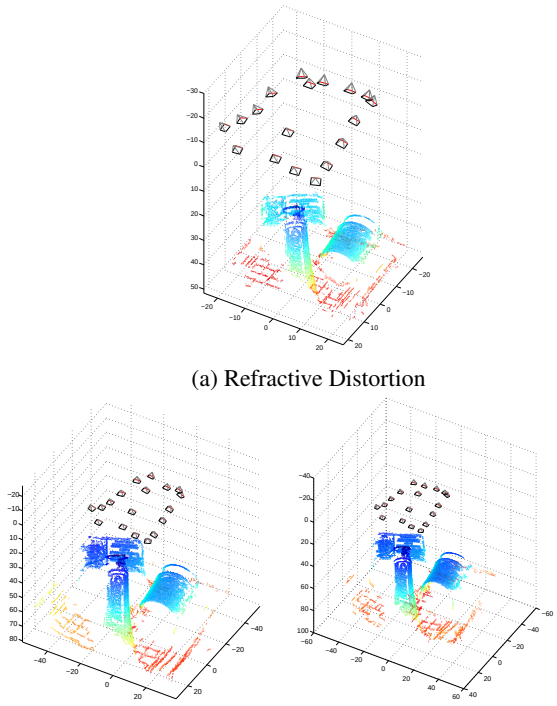
Figure 4. Sample images from the real image sequence.

modeling is larger due to the optimization is conducted in a way to minimize the reprojection error instead of the camera localization error. Even though the camera localization errors are small in all three cases, the geometry reconstruction could differ a lot.

Next we evaluate the scene reconstruction by visual inspection. The scene reconstruction of three different cases are shown in Figure 1(b)(c)(d). The reconstructed scene in the case of no distortion modeling becomes very flat and curved, while the bottom surfaces still get curved in the radial distortion case albeit its scene geometry seems to be slightly better for the two slanted plane surfaces. The reconstructed scene with our refractive distortion model faithfully recovers all three planer surfaces of the original scene geometry without any distortion.

4.2. Real data

With the success on the synthetic data, we acquire real data for qualitative comparison. A water tank is of size $15 \times 12.5 \times 11$ (L×W×H $inch$), filled with water 9 $inch$ high. A checkerboard pattern is attached to the bottom of the water tank for visual observation of the distortion pattern. Images are captured with a Google Nexus S smartphone, with the orientation readings obtained from onboard accelerometer and magnetic field sensors. Sample images are shown in Figure 4. The scene contains a rectangle coffee can and a round tea can placed on the bottom, and a toothpaste leaned on the coffee can at one end and the bottom of the water tank at the other end. In total 18 images are taken following a path around the water tank. The path is faithfully estimated by our camera localization with refractive distortion modeling as shown in Figure 5(a). Figure 5(b) and (c) show the results of scene reconstruction for no distortion model and radial distortion model, respectively. The reconstructed scene gets distorted especially for the bottom checkerboard plane for these two cases, where



(a) Refractive Distortion

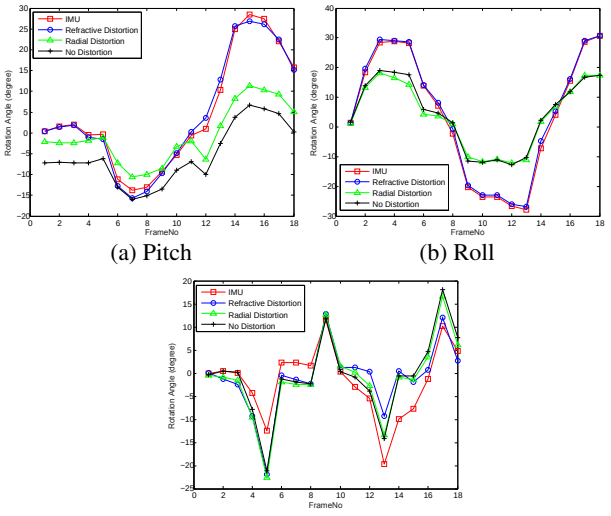
(b) No Distortion (c) Radial Distortion

Figure 5. The estimated camera path and scene reconstruction for the real image sequence. The color in the scene reconstruction encodes the height of a point from the lowest (red) to the highest (blue). (Best viewed in color.)

the geometry distortion is a little less severe with the radial distortion model. Another noticeable distortion is located at the top face of the rectangle coffee can. With the refractive distortion model, the top surface is flat with the same height, whereas the color gradient from light blue to dark blue in the other two cases indicates a slanted surface is reconstructed. Another interesting result to look at is the estimated camera pose compared to the IMU sensor reading. As depicted in Figure 6, the estimated roll and pitch angles stay close to the IMU in the refractive distortion model, indicating the IMU sensor readings provide a very good initial guess of the roll and pitch angles for relative pose and absolute pose algorithms. The magnitude of these two angles in the no distortion model and radial distortion model is smaller than the refractive distortion model. By observing the reconstructed cameras paths in Figure 5(b) and (c), it is reasonable to have smaller angles as the camera path is closer to the center of the scene. As for yaw angle, all three methods obtain quite consistent estimation while the IMU reading deviate the estimated angle by 3° to 10°.

5. Conclusions

In this work, we explicitly model the refractive distortion as a depth-dependent function for scene reconstruction. The



(a) Pitch (b) Roll

(c) Yaw

Figure 6. The comparison of estimated camera orientations with the measured camera orientations from IMU.

advantage of using a physics-based model leads to a more realistic solution for recovering both scene geometry and camera motion. By introducing the refrax ratio, we show that bundle adjustment can be performed with a much simpler form for deriving Jacobian for non-linear optimization. By incorporating the known vertical direction, a seven-point linear solution to the relative pose estimation and a closed-form two-point solution to the absolute pose estimation can be derived. These new algorithms serve as powerful ingredients for the structure from motion framework involving the refractive plane. Promising results from experiments with synthetic and real image sequences justified the superiority of the proposed refractive distortion modeling.

As of now, all formulations are based on the assumption of flat refractive geometry involving only two mediums. A more general setting with more than two mediums and non-flat refractive geometry would be a valuable direction for further investigation.

References

[1] A. Agrawal, Y. Taguchi, and S. Ramalingam. Analytical forward projection for axial non-central dioptric and catadioptric cameras. *ECCV*, 2010.

[2] A. Agrawal, Y. Taguchi, and S. Ramalingam. Beyond alhazen's problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors. *CVPR*, 2011.

[3] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? *ICCV*, 2003.

[4] V. Chari and P. Sturm. Multi-view geometry of the refractive plane. *BMVC*, 2009.

[5] Y. Ding and J. Yu. Epsilon stereo pairs. *BMVC*, 2007.

[6] Y. Ding and J. Yu. Recovering shape characteristics on near-flat specular surfaces. *CVPR*, 2010.

[7] Y. Ding, J. Yu, and P. Sturm. Multiperspective stereo matching and volumetric reconstruction. *ICCV*, 2009.

[8] A. A. Efros, V. Isler, J. Shi, and M. Visontai. Seeing through water. *NIPS*, 2004.

[9] F. Fraundorfer, P. Tanskanen, and M. Pollefeys. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. *ECCV*, 2010.

[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Analysis Machine Learning*, 32(8), 2010. http://grail.cs.washington.edu/software/pmvs.

[11] G. Glaeser and H.-P. Schrocker. Reflections on refractions. *Journal for Grometry and Graphics*, 4, 2000.

[12] G. H. Golub and C. F. V. Loan. Matrix computations. 1996.

[13] J. Höhle. Reconstruction of the underwater object. *Photogrammetric Engineering*, 37(9):949–954, 1971.

[14] M. Kalantari, A. Hashemi, F. Jung, and J.-P. Guedon. A new solution to the relative orientation problem using only 3 points and the vertical direction. *Journal of Mathematical Imaging and Vision*, 39:259–268, 2011.

[15] R. Kotowski. Phototriangulation in multi-media photogrammetry. *ISPRS*, XXVII(V):324–334, 1988.

[16] Z. Kukelova, M. Bujnak, and T. Pajdla. Closed-form solutions to the minimal absolute pose problems with known vertical direction. *ACCV*, 2010.

[17] R. Li, H. Li, W. Zou, R. Smith, and T. Curran. Quantitative photogrammetric analysis of digital underwater video imagery. *IEEE Journal of Oceanic Engineering*, 22(2):364–375, Apr. 1997.

[18] M. A. Lourakis and A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009.

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

[20] N. J. W. Morris and K. N. Kutulakos. Dynamic refraction stereo. *ICCV*, 2005.

[21] H. Murase. Surface shape reconstruction of a nonrigid transparent object using refractino and mtion. *IEEE Trans. Pattern Analysis Machine Learning*, 14(10):1045–1052, 1992.

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

[23] M. Shortis and H. A. Beyer. Sensor technology for close range photogrammetryand machine vision. *Close Range Photogrammetry and Machine Vision*, pages 106–155, 1996.

[24] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM SIGGRAPH*, pages 835–846, 2006.

[25] Y. Tian and S. G. Narashimhan. Seeing through water: image restoration using model-based tracking. *ICCV*, 2009.

[26] T. Treibitz, Y. Y. Schechner, and H. Singh. Flat refractive geometry. *CVPR*, 2008.

[27] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. *LNCS*, 1883:153–177, 2000.