Refractive Structure-from-Motion Through a Flat Refractive Interface

François Chadebecq[1,2] Francisco Vasconcelos[1,2] George Dwyer[1,2] René Lacher[1]
Sébastien Ourselin[2] Tom Vercauteren[2] Danail Stoyanov[1,2]
[1]Surgical Robot Vision, UCL London, UK
[2]Centre for Medical Image Computing, UCL, London, UK
f.chadebeq@ucl.ac.uk

Abstract

Recovering 3D scene geometry from underwater images involves the Refractive Structure-from-Motion (RSfM) problem, where the image distortions caused by light refraction at the interface between different propagation media invalidates the single view point assumption. Direct use of the pinhole camera model in RSfM leads to inaccurate camera pose estimation and consequently drift. RSfM methods have been thoroughly studied for the case of a thick glass interface that assumes two refractive interfaces between the camera and the viewed scene. On the other hand, when the camera lens is in direct contact with the water, there is only one refractive interface. By explicitly considering a refractive interface, we develop a succinct derivation of the refractive fundamental matrix in the form of the generalised epipolar constraint for an axial camera. We use the refractive fundamental matrix to refine initial pose estimates obtained by assuming the pinhole model. This strategy allows us to robustly estimate underwater camera poses, where other methods suffer from poor noise-sensitivity. We also formulate a new four view constraint enforcing camera pose consistency along a video which leads us to a novel RSfM framework. For validation we use synthetic data to show the numerical properties of our method and we provide results on real data to demonstrate performance within laboratory settings and for applications in endoscopy.

1. Introduction

A variety of underwater activities rely on video imaging and can be supported by computer vision methods for mapping the environment for enhanced navigation and exploration. Recovering the 3D geometry of the scene and the motion of the camera requires adaptation to the multi-view geometry used for reconstruction in air. To formulate Refractive Structure-from-Motion (RSfM) it is important to consider the deviation from the classical pinhole camera model used to describe image formation (see Figure 1).

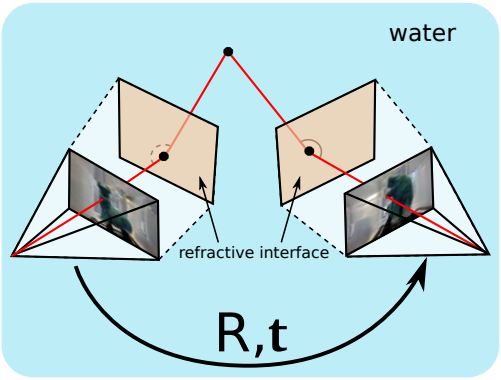

Figure 1. Refractive Structure-from-Motion for underwater imaging assuming a thin refractive plane.

Light rays undergo refraction when passing through mediums with different optical density as defined by Snell's law. The angle of deflection depends on the refractive index of traversed medium as well as the incidence angle of the incoming light ray. This has a strong influence on the image formation process as it invalidates the single view point assumption [10]. While adapting intrinsic camera parameters and distortion coefficients can compensate for refraction it introduces a systematic geometric bias which affects 3D measurements and camera pose estimation [18, 16, 24] (see Figure 2 *(a)*). Moreover, if advanced radial basis distortion functions (RBF) are able to compensate for severe and irregular distortions [1, 30], refractive distortion directly depends on the depth of the 3D points in the scene [27]. Therefore, RBF models only provide a reasonable approximation within a limited range of distances.

It has been demonstrated that vision through a refractive interface can be modelled by an axial camera to avoid such bias, however, ray-based models are difficult to calibrate due to the high dimensionality of their parametrisations [10, 2]. Even if calibrated well, 3D reconstruction with a moving general camera remains a challenge underwater. Although the generalised camera model encompasses the

axial camera model, pose estimation is highly sensitive to noise and computationally unstable. This is particularly evident for monocular axial camera model for which conditioning of light rays is critical. Approaches to formulate the problem typically make prior assumptions, such as knowledge of the camera orientation, or considering a camera moving behind a fixed refractive interface (such as a camera looking through an aquarium) which is not suited to an immersed camera moving underwater [5, 4].

In contrast, refractive camera models explicitly consider one or several parallel interfaces separating the optical system from one or multiple mediums with different refractive indexes [2] (see Figure 2 *(b)*). Approaches for 3D reconstruction relying on refractive geometric constraints have been reported especially for the case of a watertight shielded camera casing [14, 15, 13]. More particularly, they focus on deep underwater imaging and thus consider a thick refractive interface (see Figure 2 *(c)*). The two-view relative pose problem can be iteratively solved using geometric constraints [2] followed by bundle adjustment by associating to each 2D point a virtual perspective camera. This formulation does not rely on the refractive re-projection error, which would required solving a 12^{th} degree polynomial [29]). Dense 3D reconstruction is obtained by using a refractive plane sweep algorithm [14]. However, this method requires a good initialisation and addresses the case of a thick glass interface implying severe refractive distortion effects. All methods for estimating relative camera motion are particularly sensitive to noise. this is a major limitation because underwater images are subject to complex light scattering and diffusion, as well as medium turbidity, which make precise point constraints difficult to establish.

Contribution: We propose a novel RSfM framework for a camera looking through a thin refractive interface with the following theoretical developments:

- Formulation of a new four-view constraint derived from the refractive geometry, which is important for relative pose estimation consistency over consecutive video frames.

- A new RSfM pipeline that relies on the the refractive fundamental matrix derived from the generalised epipolar constraint [22], which we use with refractive re-projection constraints to refine an initial estimate of the relative camera pose estimated using the adapted pinhole model with lens distortion [18].

The proposed method applies to underwater imaging scenario where camera's lens is directly in contact with water (e.g. endoscopic surgery such as arthoscopy, consumer action camera, see Figure 1).

We succinctly review previous work in Section 2. In Section [5] we recall single view refractive geometry, and then in Section 3 we derive the two-view refractive geometry that leads into the formulation of the refractive fundamental matrix and a novel four-view refractive constraint. Section 4 describes the complete RSfM pipeline. On Section 5 we demonstrate the improvements on numerical stability that our new approach brings by presenting results on both synthetic and real data.

Notation: The world reference frame (X, Y, Z) is arbitrarily set for all viewpoints. The Z-axis lies on the camera axis defined as the line passing through the normal of the refractive interface ($\mathbf{n} = (0\ 0\ 1)^\top$) and the camera optical centre. The X and Y axes lie on the refractive plane and respectively align with the X_c-axis and Y_c-axis of the camera coordinate frame. The pose of the camera is expressed as $\mathsf{P}_p = \mathsf{R}_p^{-1}(\mathsf{I} - \mathbf{t}_p)$ where R_p corresponds to the refractive plane orientation relative to the camera coordinate frame and $\mathbf{t}_p = (0\ 0\ d)^\top$. The interface to camera centre distance along the camera's axis is denoted as d. An image point i observed in view j is denoted $\mathbf{p}_j^i = (x\ y\ 1)^\top$. We denote $\mathbf{P}_j^i = (x\ y\ z\ 1)^\top$ as the point of incidence (point lying on the refractive interface) related to the 3D point $\mathbf{Q}^i$ projected in $\mathbf{p}_j^i$. The corresponding refracted light ray (i.e. travelling within the water tightness housing) is expressed by $\mathbf{q}_j^i = (q_{j,x}^i\ q_{j,y}^i\ q_{j,z}^i)^\top = ((\mathsf{R}_r^{-1}\tilde{\mathbf{p}}_j^i)^\top\ 0)^\top$ where $\tilde{\mathbf{p}}_j^{i\top}$ is the unit vector corresponding to the image point $\mathbf{p}_j^i$.

Light rays are defined by a starting point (e.g. a point of incidence) and a direction vector denoted $\mathcal{L}$. The Plücker coordinates of a light ray are denoted $\mathbf{L} = (\mathbf{L}_0, \ldots, \mathbf{L}_6)^\top$ [27]. As such, $\mathbf{L}_{(a,b,c)}$ defines a vector composed by the elements $\mathbf{L}_a$, $\mathbf{L}_b$ and $\mathbf{L}_c$ of $\mathbf{L}$. The vector $\hat{\mathbf{v}} = (v_x{}^2\ v_xv_y\ v_y{}^2\ v_xv_z\ v_yv_z\ v_z{}^2)^\top$ denotes the lifted coordinate of the 3D vector $\mathbf{v}$ and hence if two vectors are related by a linear transformation T such as $\mathbf{v_1} = \mathsf{T}\mathbf{v_2}$, their lifted coordinates are related by $\hat{\mathbf{v}}_\mathbf{1} = \mathsf{D}_s^{-1}\mathsf{S}(\mathsf{T} \otimes \mathsf{T})\mathsf{S}^\top\hat{\mathbf{v}}_\mathbf{2}$. The symbol $\otimes$ refers to the Krönecker product and the two design matrix D_s and S are defined as $\mathsf{D}_s = diag(1\ 2\ 1\ 2\ 2\ 1)$ and $\mathsf{S}([1,1],[2,2],[2,4],[3,5],[4,3],[4,7],[5,6],[5,8],[6,9]) = 1$.

2. Prior Work on Underwater SfM

An exhaustive survey on underwater 3D reconstruction methods can be found in [21]. The majority of approaches for underwater 3D reconstruction rely on standard SfM methods assuming the adapted pinhole camera model [12, 16], however, this often leads to inaccurate 3D reconstruction [20]. Additionally, systematic geometric bias is present [24] and when the refractive interface is not fronto-parallel to the image plane, measurement errors are particularly significant [15]. Relying on a ray-based model and considering a camera moving behind a fixed refractive plane allows derivation of the refractive fundamental matrix relationship [5]. However, it is defined by a 15×15 matrix and as a result estimation is computationally unstable due to the

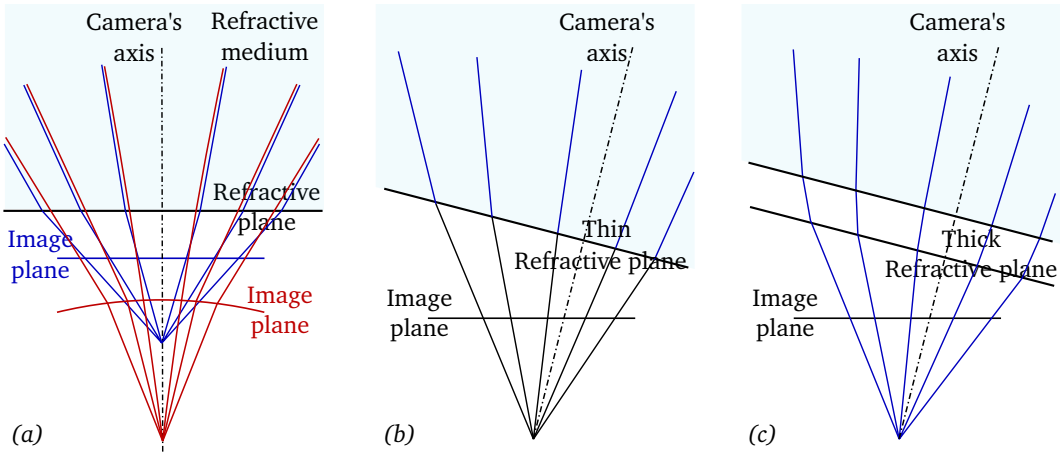

Figure 2. Underwater image formation model: *(a)* Structure-from-Motion methods for underwater imaging assume the pinhole camera model compensating for refraction effect by adapting focal length and distortion parameters (red). Refractive camera model explicitly consider refraction due to a change of medium refractive index (blue). *(b)* Refractive camera model considered in our study [2] for modelling thin glass interface. The orientation of the refractive plane has a significant influence on the image formation process. *(c)* RSfM method for deep underwater imaging [13] assume a thick refractive plane implying a significant refractive distortion effect.

many degrees of freedom. An alternative ray-based model allows the underlying refractive geometry to be expressed as a direct extension of the projective geometry but this only allows 3D reconstruction are obtained up to a similarity and assumes that refraction occurs at the camera centre [6].

Modelling the refractive interface leads to the explicit RSfM formulation in the case of a fixed interface [4]. The method leads to promising results but requires camera's motion to be partly known, for example thanks to an additional sensor such as an intertial measurement unit (IMU). Assuming a stereo rig and camera rotation is known, [17] provide with an optimal solution to the relative translation problem under L_∞ norm. This method can be extended to unknown rotations assuming a thin refractive plane parallel to both image plane of the cameras. More recently, [11] developed efficient minimal solvers for absolute camera pose estimation under a fixed refractive interface. A complete RSfM framework for the case of a camera embedded in a watertight case has been derived in [15, 14, 13]. Relative camera motion between two successive views is estimated by relying on the flat refraction and co-planarity constraints [2] following a non-linear refractive bundle adjustment extending a previous formulation [23]. Dense depth estimation is obtained by using a refractive plane sweep algorithm [14] that relies on near optimal initialisation.

RSfM has also been considered to solve for the absolute scale ambiguity inherent to SfM [25]. Knowing the position and orientation of the interfaces theoretically yields the absolute camera motion as relative pose is no longer invariant to scale change in camera translation [15]. However, this is particularly sensible to noise and only considered assuming a thick refractive interface. It has been experimentally observed that RSfM methods cannot reliably infer the absolute scale of a scene even for thick refractive interface considered in deep underwater imaging [13].

3. Refractive Multiple View Geometry

3.1. Refractive Camera Model

A detailed survey of underwater camera models is available in [24]. We explicitly consider refraction at an interface as developed by [2] who showed that the refractive camera model corresponds to an axial camera. By formulating refractive constraints on the plane of refraction (onto which will lie the camera axis and an incident light ray), they propose a direct method for calibrating the position and orientation of one or multiple refractive interfaces. We recall the so-called co-planarity constraints which has led the authors to derive the refractive forward projection equation that we will use throughout this paper (see Figure 3). It enforces each light ray to lie on the plane of refraction and the refracted ray to intersect the camera axis. This is mathematically described by:

$$(\mathrm{R}\mathbf{Q}^i + \mathbf{t})^\top (\mathbf{n} \times \mathbf{P}^i) = 0 \tag{1}$$

Considering a single refractive interface, the co-planarity constraint can be developed leading to the refractive forward projection function. It is expressed by a 4th degree polynomial equation:

$$(\mathbf{Q}^i_{px} - \mathbf{P}^i_{px})^2 (d^2\mu_2^2 + \mu_2^2 {\mathbf{P}^i_{px}}^2) - (d\mathbf{P}^i_{px} - \mathbf{Q}^i_{py}\mathbf{P}^i_{px})^2 = 0 \tag{2}$$

where d corresponds to the distance from the camera's optical center to the refractive plane and μ_2 corresponds to the refractive index of the external medium. The axis z_1 aligns

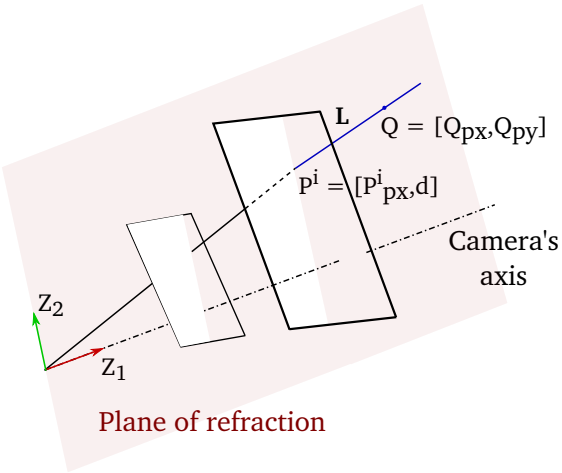

Figure 3. The refractive forward projection equation is defined over the plane of refraction. It is defined by an incident of refracted light ray and the camera's axis [2].

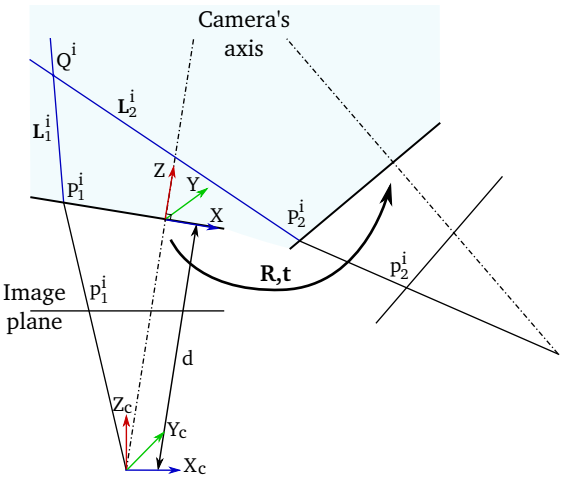

Figure 4. Two-view refractive geometry assuming a thin glass interface. This model applies to camera embedded within a thin watertight case or whose lens is in direct contact with the water.

with the camera axis, and $z_2 = z_1 \times (z_1 \times \mathbf{Q}^i)$ defines the orthogonal coordinate frame $[z_2, z_1]$ of the plane of refraction. The point $\mathbf{Q}^i = [\mathbf{Q}^i_{px}, \mathbf{Q}^i_{py}]$ expresses the 3D point $\mathbf{Q}^i$ in this coordinate frame. The refracted light ray is defined by $\mathbf{P}^i_{px} z_2 + dz_1$ where $\mathbf{P}^i_{px}$ corresponds the unknown projection depth parameter. We will refer to this refractive projection function as P_r throughout the paper. For the sake of clarity, we first remind the single-view refractive geometry introduced in [5] although we will consider the forward refractive reprojection equation derived in [2] (see equation 2). We then formulate the two-view refractive fundamental relationship in the form of the generalized epipolar constraint for axial cameras (see Figure 4). We finally derive a novel four-view constraint in the last subsection.

3.2. Single-View Refractive Geometry

The refractive point $\mathbf{P}^i$ can be expressed by:

$$\mathbf{P}^i = (-d\frac{q^i_x}{q^i_z} \; -d\frac{q^i_y}{q^i_z} \; 0 \; 1)^\top \tag{3}$$

According to Snell's law, the corresponding incident ray (i.e. running through the refractive medium) is defined by:

$$\mathcal{L}^i = (\lambda q^i_x \; \lambda q^i_y \; \sqrt{1 - \lambda^2 + \lambda^2 {q^i_z}^2} \; 0)^\top \tag{4}$$

where λ refers to the external medium refractive index.

Using Plücker coordinates, $\mathbf{L}^i$ can be reformulated as:

$$\mathbf{L}^i = (\lambda q^i_x \; \lambda q^i_y \; v^i_z - d\frac{q^i_y}{q^i_z} v^i_z \; d\frac{q^i_x}{q^i_z} v^i_z \; 0)^\top \tag{5}$$

where $v^i_z = \sqrt{1 - \lambda^2 + \lambda^2 {q^i_z}^2}$.

The assumption of a line $\mathbf{L}^i_1$ intersecting the incident ray $\mathbf{L}^i$ is verified by the Klein constraint [27]:

$$\mathbf{L}^i_1 \mathsf{W} \mathbf{L}^i = 0 \tag{6}$$

where $\mathsf{W}_{6\times6} = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}$. It thus allows [5] to formulate the following refractive projection equation:

$$\left(\widehat{\mathbf{L}}^\top_{1^i_{(6,1,2)}} \; \widehat{\mathbf{L}}^\top_{1^i_{(4,5,3)}}\right) {}_r\mathsf{P}^\top (\frac{\hat{\mathbf{q}^i}}{{q^i_z}^2} \; \hat{\mathbf{q}^i})^\top = 0 \tag{7}$$

The refractive projection matrix ${}_r\mathsf{P}$ is defined as:

$${}_r\mathsf{P} = \mathsf{D}_s^\top \begin{pmatrix} (1-\lambda^2)\mathsf{D}_s^{-1}\mathsf{S}_s\mathsf{t}_s^\top \otimes \mathsf{t}_s^\top\mathsf{S}_s^\top & 0 \\ \lambda^2\mathsf{D}_s^{-1}\mathsf{S}_s\mathsf{t}_s^\top \otimes \mathsf{t}_s^\top\mathsf{S}_s^\top & -\lambda^2\mathsf{D}_s^{-1}\mathsf{S}_s\mathsf{t}_t^\top \otimes \mathsf{t}_t^\top\mathsf{S}_s^\top \end{pmatrix} \tag{8}$$

where $\mathsf{t}_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and $\mathsf{t}_s = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -d & 0 \\ d & 0 & 0 \end{pmatrix}$.

The refractive projection matrix is of size 12×12 and it expresses the refractive projection of a 3D line onto a quartic curve in the image plane. Alternatively, the refractive forward projection function derived in [2] projects a 3D point onto the corresponding refractive point $\mathbf{P}^i$.

3.3. Two-view Refractive Geometry

We now consider the incident ray $\mathbf{L}_2$ giving rise to the point $\mathbf{p}^i_2$ in the second view. The ray $\mathbf{L}^i_2$ is defined by:

$$\mathbf{L}^i_2 = \mathsf{T}(\lambda q^i_{2,x} \; \lambda q^i_{2,y} \; v^i_{2,z} \; -d\frac{q^i_{2,y}}{q^i_{2,z}} v^i_{2,z} \; d\frac{q^i_{2,x}}{q^i 2,z} v^i_{j,z} \; 0)^\top \tag{9}$$

where $\mathsf{T} = \begin{pmatrix} \mathsf{R} & 0 \\ [t]_x\mathsf{R} & \mathsf{R} \end{pmatrix}$, $[t]_x = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix}$ and $v^i_{j,z} = \sqrt{1 - \lambda^2 + \lambda^2 {q^i_{j,z}}^2}$.

The refractive two-view relationship can therefore be explicitly formulated in the form of the generalised epipolar constraint [22, 28]. Relying on the Klein constraint 6, the refractive fundamental constraint can be defined by:

$$\begin{bmatrix} \lambda q^i_{1,x} \\ \lambda q^i_{1,y} \\ v^i_{1,z} \\ -d\frac{q^i_{1,y}}{q^j_{1,z}} v^i_{1,z} \\ d\frac{q^i_{1,x}}{q^i_{1,z}} v^i_{1,z} \end{bmatrix}^\top \underbrace{\begin{pmatrix} & & & R_{11} & R_{12} \\ & [t]_x \mathsf{R} & & R_{21} & R_{22} \\ & & & R_{31} & R_{32} \\ R_{11} & R_{12} & R_{13} & 0 & 0 \\ R_{21} & R_{22} & R_{23} & 0 & 0 \end{pmatrix}}_{\mathsf{rF}} \begin{bmatrix} \lambda q^i_{2,x} \\ \lambda q^i_{2,y} \\ v^i_{2,z} \\ -d\frac{q^i_{2,y}}{q^i_{2,z}} v^i_{2,z} \\ d\frac{q^i_{2,x}}{q^i_{2,z}} v^i_{2,z} \end{bmatrix} \tag{10}$$

The generalised relative pose problem can be estimated using a minimal number of 6 points correspondences [26]. It is however particularly noise-sensitive. Although it can be used within a robust estimation framework, it remains unsuitable to underwater scenario where feature matching is critical. Moreover, normalizing features vectors (incident light ray) in the case of a monocular axial camera is complex. Expressing equation 10 in the form of a norm-constrained homogeneous linear least squares leads to ill-conditioned and rank-deficient feature matrix. Alternatively, a linear and effective algorithm using a minimum of 16 points correspondences has been proposed in [19]. The authors propose an iterative method where first the rotation component is estimated from $\mathsf{E} = [t]_x \mathsf{R}$, and then the translation component is extracted from 10. The translation component is theoretically estimated without scale ambiguity. We however observed that such approaches cannot be efficiently adapted to the case of underwater vision.

These observations suggest a two-step approach where relative camera poses are first estimated assuming the adapted pinhole camera model and then refined relying on refractive reprojection constraints 2 as well as the refractive fundamental constraint 10. The first step of the proposed approach provides with a reasonable estimates of camera poses but also an effective way to discard outlier correspondences. We have distinguished two cases for camera pose refinement. Assuming wide-baseline camera motion and the refractive interface parallel to the image plane, we observed that the adapted pinhole model provide with accurate camera rotation estimation while translation is significantly affected by refractive distortion (i.e. only minimizing to $\mathbf{t}$, see equation 11). When the refractive interface is tilted and for small-baseline camera motion, SfM is particularly sensitive to both noise and refractive effect. Therefore, we refine for both rotation and translation leading to the following non-linear constraint:

$$\begin{aligned} \arg\min_{\theta,\mathbf{t}} \sum_{i=1}^{N} \|P_p P_r^1 Q^i - q_1^i\|^2 + \sum_{i=1}^{N} \|P_p P_r^2 Q^i - q_2^i\|^2 \\ + \sum_{i=1}^{N} \|\mathbf{L}_1^i \mathsf{rF} \mathbf{L}_2^i\|^2 \end{aligned} \tag{11}$$

The choice of two different strategies can be explained by the weak robustness of SfM considering small baseline camera motion in underwater imaging. On the other side, for wide-baseline camera motion, even a small amount of noise introduces camera pose ambiguities as it is generally compensated by slight camera orientation adjustments due to the non-convex nature of geometric refractive constraints. The proposed method is summarized in algorithm 1.

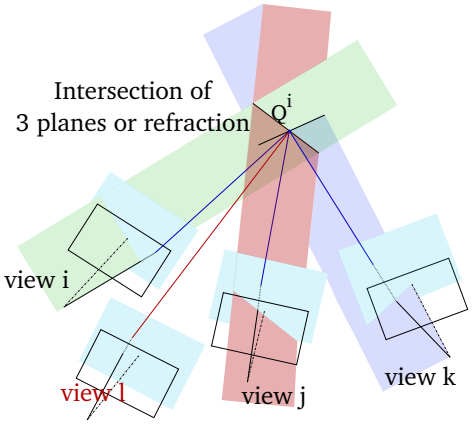

Figure 5. Four-view refractive geometry relationship. Considering the intersection of three planes of refraction i, j and k defined by the light rays corresponding to the 3D point $\mathbf{Q}^i$ and the camera's axis, we refine estimates of the forth view camera pose by ensuring light ray l to pass through $\mathbf{Q}^i$.

Algorithm 1: Two-view Refractive SfM

Require: pinhole camera parameters, position and orientation of the refractive interface, a pair of consecutive views.

1: Extract and match SIFT features
2: Estimate camera motion using adapted pinhole camera parameters [18, 12]
3: Discard mismatches based on reprojection error threshold (1 pixel in our experiments)
4: Undistort image point using pinhole camera parameters
5: Refine camera pose estimates by minimizing 11

3.4. Four-view Refractive Geometry

In order to enforce camera pose consistency along a video sequence, we propose a novel refractive constraint assuming 3D points are visible in four successive views (see Figure 5). Considering a general camera motion, a 3D point Q^i can be expressed as the intersection of three planes of refraction. There are six degenerate cases for which the constraint cannot be applied: at least two planes of refraction are parallel or coincident, the three planes of refraction intersect in a line and each plane cuts the other two in a line. In such cases, the intersection of the three planes is either a plane, a line or does not exist. Such degenerate cases rarely appear in practice, unless working in highly planar environments, as refractive distortion depends on the depth of the 3D scene. Degenerate cases can also be efficiently detected and thereafter discarded by inspecting the rank of the coefficient and augmented matrices derived from the equations of the planes of refraction [12]. We therefore formulate a constraint enforcing the corresponding incident light ray in the fourth view to pass through the point Q^i. As

such, a set of four views gives raise to three constraints that we mathematically express by the following expression:

$$\underset{\mathbf{T_i},\mathbf{T_j},\mathbf{T_k},\mathbf{T_l}}{\arg\min} \sum_{i,j,k=\binom{4}{3},l} \| \mathbf{T_l}\mathbf{L}_{l(4,5,6)} - \left(\tilde{\mathbf{Q}}^i \times \mathbf{T_l}\mathbf{L}_{l(1,2,3)} \right) \|^2 \tag{12}$$

where $\tilde{\mathbf{Q}}^i$ corresponds to the intersection of the planes of refraction of the views i, j and k and $\mathbf{T_i},\dots\mathbf{T_l}$ defines the linear transformation corresponding to camera's poses. The point of intersection $\tilde{\mathbf{Q}}^i$ is computed by solving the system of linear equations defining each of the plane of refraction (expressed through their Hessian form, $\vec{n}_i = \mathbf{R}_i \left(\mathbf{L}_{(1,2,3)} \times \mathbf{n}\right), \mathbf{R_i} \left(\mathbf{L}_{(1,2,3)} \times \mathbf{n}\right) \mathtt{t}_i$).

The refractive four-view constraint does not depend on the refractive index of the external medium unlike the refractive fundamental constraint 10. As such this constraint can be easily extended to the complex case of multiple refractive interfaces. Unlike the classical refractive bundle adjustment approach, which suffers from a high computational cost, our method is particularly efficient providing a direct solution.

4. Multiple-view Refractive Structure-from-Motion

Algorithm 2: Multiple-view Refractive SfM

Require: pinhole camera parameters, position and orientation of the refractive interface, video sequence

Optional: Reference scale

1: Estimate structure and motion for the first pair of views using algorithm 1
2: Rescale camera pose according to the reference scale
3: **for** each new frame of the video sequence
4: Extract and match SIFT features in views i and $i-1$
5: Estimate camera motion using adapted pinhole camera parameters [9]
6: Discard mismatches based on reprojection error threshold (1 pixels in our experiments)
7: Undistort image point using pinhole camera parameters
8: Refine camera pose estimates by minimizing 11
9: **if** $i > 4$ **then**
10: Use four-view constraint 12 to enforce camera motion consistency

The proposed RSfM framework is summarized by algorithm 2. We follow a strategy similar the one presented in algorithm 1. We first solve for the perspective-n-point problem assuming adapted pinhole camera model [9] and refine camera pose assuming the refractive camera model. As previously mentioned, absolute camera motion cannot be accurately estimated even for very low noise. This is more particularly the case considering vision through a flat refractive interface.

5. Experiments

The proposed RSfM method has been evaluated on both synthetic and real datas. For the synthetic experiments, we consider underwater scenarios where the scene is imaged at a distance between 3 and 4 meters by a consumer action camera. For real experiments, we first consider a similar scenario but for a scene situated at a distance of approximately 500 mm. We then highlight a particular application for RSfM and show results for endoscopy.

5.1. Synthetic Data

The synthetic dataset has been generated by considering the following setup. We assumed a consumer action camera whose focal length is 800 pixels and resolution capture is 1280×960. The position of the refractive interface has been randomly chosen between 3 and 50 mm. We assumed the interface is either fronto-parallel to the image plane or it has been shifted at an angle of 15 degrees (along a random axis). The camera observed a 3D point cloud (200 points) randomly generated within a cube of size 1 meter and at a distance between 3 and 4 meters. We considered the camera motion follows a curvilinear path but we randomly generated camera poses along this path as well as camera to scene distances. We relied on the forward refractive projection function (equation 2) and considered underwater scenarios ($\lambda = 1.3$).We added a Gaussian noise of 1 pixel standard deviation (std) to virtual image points. We compared our method with SfM assuming adapted pinhole camera parameters. As such we generated a set of synthetic calibration images (without image noise) in order to estimate adapted pinhole parameters. As expected, when the interface is fronto-parallel to the image plane, adapted focal length is to one decimal place equal to $1.3 * f = 1040$. We furthermore considered the 6^{th} degree Brown-Conrady model for distortion [3].

We first report results on camera motion estimation for two successive frames of a video sequence. We considered 100 views randomly selected along an ∞-shaped curve path and compared our method with SfM assuming adapted pinhole camera model. Results reported in Figure 6 were obtained for small-baseline camera motion while results presented in Figure 7 were obtained for wide-baseline camera motion. For both of these figures, the top row corresponds to the results obtained considering the refractive interface is parallel to the image plane. The bottom row corresponds to the results obtained when the refractive interface is set at an angle of 15 degrees. We observed that the proposed RSfM method significantly improves initial pinhole estimates for both translation and rotation. It is more significant for small-baseline camera motion or when the refractive interface is tilted. In this case, we have not been able to provide with significant results using SfM while RSfM allows us to obtain accurate 3D reconstruction despite important level of noise. This explains the constant motion estimation error. Moreover, RSfM allow us to efficiently estimates camera poses even for small-baseline camera motion despite a greater sensitivity towards noise. These results validate the two different refinement strategies defined for wide-baseline camera motion when the refractive interface is parallel to the image plane.

We then report results for estimation of camera trajectory for 40 frames. Camera poses were randomly generated along a curvilinear path. We present in Figure 8 results obtained assuming the refractive interface is parallel to the image plane (left) or that it has

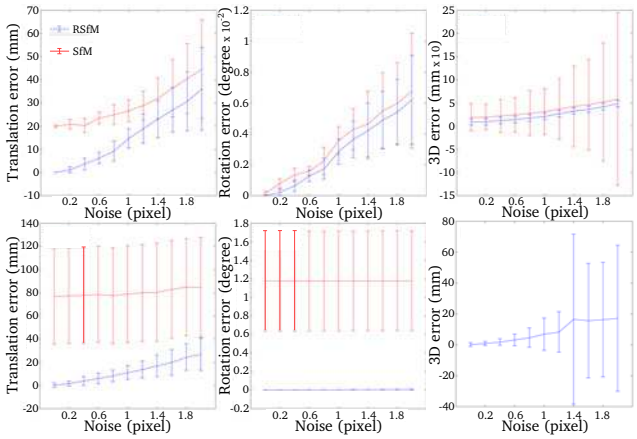

Figure 6. Synthetic evaluation of camera pose estimation for two successive views of a video sequence and small-baseline camera motion. *Top row:* no interface tilt. *Bottom row:* interface tilted at an angle of 15 degrees.

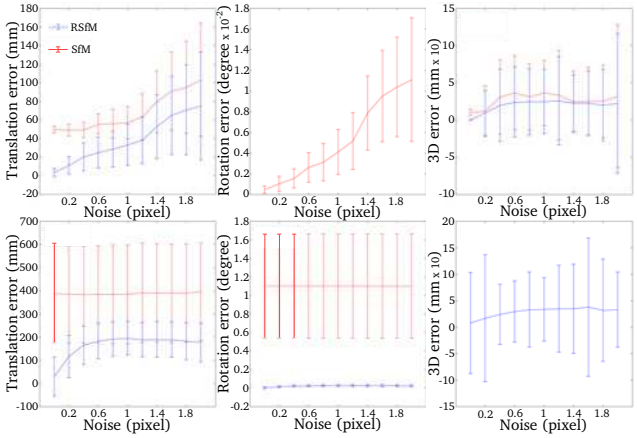

Figure 7. Synthetic evaluation of camera pose estimation for two successive views of a video sequence and wide-baseline camera motion. *Top row:* no interface tilt. *Bottom row:* interface tilted at an angle of 15 degrees.

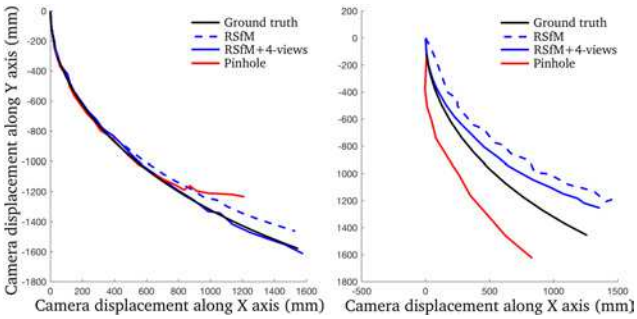

Figure 8. Synthetic evaluation of camera pose estimations for 50 views along a curvilinear path. *Left*: no interface tilt, image noise of 1 pixel. *Right*: interface tilted at an angle of 15 degrees, image noise of 1 pixel.

been tilted at an angle of 15 degrees (right). Results demonstrate the effectiveness of the four-view refractive constraint which enforces camera trajectory consistency along a video. This is more particularly the case when the refractive interface is parallel to the image plane. In such case, we observed that SfM pose estimation drift. When the refractive interface is not parallel to the image plane, we observed that the four-view constraint corrected for camera pose drifting despite SfM suffered from a significant drift. We however noticed that it will be necessary to enforce the global consistency of camera trajectory using bundle adjustment.

5.2. Real Data

We evaluated the effectiveness of the proposed RSfM framework in a laboratory environment. We first compare SfM and RSfM for two-view relative pose estimation. For this purpose, we used a stereo endoscope for which absolute camera pose was known. We then consider two kind of optical equipments; a consumer action camera and a medical endoscope. We more particularly highlight underwater 3D reconstruction accuracy and compare 3D shape estimation obtained with both SfM and RSfM. Ground truth was obtained using an Artec Spider 3D scanner (Artec 3D®). The different point clouds were aligned with the ground truth mesh using Iterative Closest Point [31]. Discrepancy measurements were computed as the minimal distance between each point cloud and the reference mesh.

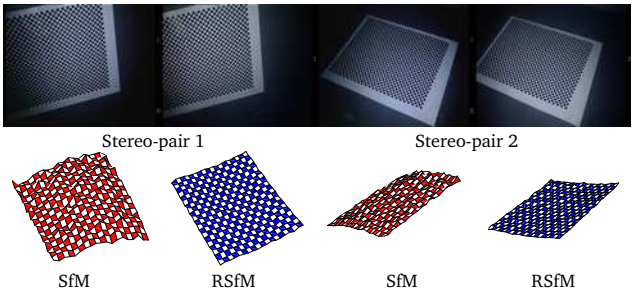

Figure 9. Underwater 3D reconstruction of a checkerboard pattern using a surgical stereo endoscope. RSfM significantly improves the 3D shape estimated by SfM.

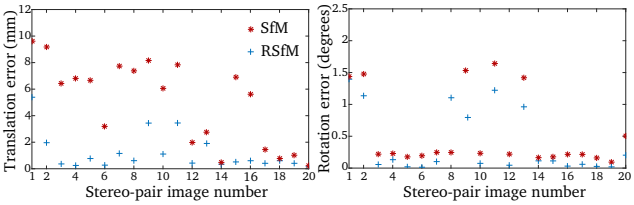

Figure 10. Relative pose estimation between the cameras pair of a surgical stereo-endoscope. Despite the small camera baseline, SfM estimations are efficiently refined by our RSfM method.

Checkerboard dataset: For reliable quantitative and qualitative analysis of camera pose estimation, we performed experiments using a stereo camera observing a planar checkerboard underwater (20 poses). We estimated relative camera pose between the stereo pair using both two-view SfM and RSfM. We considered as the ground truth the rigid pose estimated by calibrating the endoscope in air (translation: 5.8 mm, rotation: 3.5 degrees along the Y

axis). We noticed a mean rotation error of 0.54 degrees with a std of 0.58 degrees for SfM while it is of 0.35 degrees with a std of 0.51 degrees for RSfM. More importantly, RSfM significantly improved translation estimation with an error of 5.24 mm with a std of 3.21 mm for SfM and 1.21 mm with a std of 1.39 mm for RSfM. Results are presented in Figures 9 and 10.

Hippopotamus dataset: We reproduced in a lab environment the imaging conditions corresponding to a consumer action camera imaging a scene situated at a distance of approximately 500 mm. We used a go pro camera® that we immersed within a tank filled with water. We imaged a statuette (of size approximately 150 $\times$ 130 mm) that we manually rotated in front of the camera (10 views). It is worth to note that the consistent calibration of such cameras underwater and in air is complex due to their optical properties (e.g. wide field of view). It thus introduces an additional bias affecting camera pose estimation.

We present in Figure 11 a close look at the 3D reconstruction obtained using the proposed RSFM approach and four views of the hippopotamus statuette. It highlights the accuracy of the proposed method as well as its robustness toward noise. The wide field of view of the action camera used for this experiments prevents accurate visual odometry; nevertheless, we compared 3D reconstruction result using SfM and RSfM considering two consecutive views of the statuette. We observed a root mean square error of 5.3 mm with a std of 3.6 mm for SfM while we obtained an error of 4.6 mm with a std of 3.2 mm for RSfM.

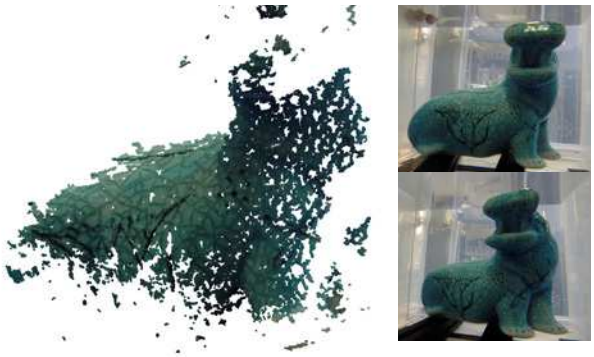

Figure 11. Close look at the underwater 3D reconstruction of a statuette using the proposed RSfM approach.

Rabbit dataset: We evaluated our RSfM framework considering fluid-immersed endoscopic imaging. Using a setup similar to the one described for the hippopotamus dataset we immersed a small toy (of size approximately 15 $\times$ 15 mm) within a tank filled with water. Images have been acquired at a distance approximately between 30 and 80 mm which corresponds to the working distance of the endoscopic equipment used for our experiments (15 views). An illustration of the achieved results is presented in Figure 12.

Due to the small-baseline camera motion we have not been able to provide with reliable results using classical multiple-view SfM methods. Despite the lack of ground truth, we observed that camera poses estimated by our RSfM method correspond to the manual displacement of the camera around the toy. We moreover observed that we recover the 3D shape of the toy despite small-baseline camera motions which validates the effectiveness

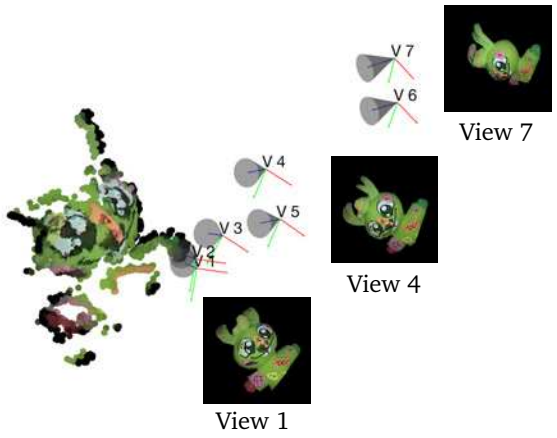

Figure 12. Underwater 3D reconstruction of a toy using the proposed RSfM approach. Results have been obtained by acquiring seven images ($V_1, \ldots, V_7$, among which images corresponding to views 1, 4 and 7) using a medical endoscope.

our synthetic experiments and the validity of the proposed approach. Comparing SfM and RSfM using two consecutive views of the toy, we observed a similar 3D reconstruction error of respectively 0.3 mm with a std of 0.4 mm and 0.2 mm with a std of 0.5 mm. We nevertheless observed a significant improvement in the uniformity of shape. More experiments are needed in order to evaluate its applicability to complex fluid-immersed scenario.

6. Conclusion

We proposed a novel RSfM framework for underwater 3D reconstruction and camera motion estimation. We more particularly address the case of cameras for which sealing of the optical system is ensured by a thin glass interface. We succinctly derived the refractive fundamental matrix and combined it with the refractive re-projection error to refine pose estimates obtained by assuming the pinhole model. We also derived a novel four view constraint allowing us to enforce camera motion consistency along a video. We evaluated the proposed RSfM framework on both synthetic and real data and demonstrated its efficiency toward SfM generally considered for underwater 3D reconstruction.

A perspective work will be to integrate the proposed approach within underwater mosaicking [7] or Refractive Simultaneous Localization And Mapping pipeline. This will require to develop robust underwater registration methods adapted to intended applicative context (e.g learning-based approaches) or combine it with robotic imaging [8]. It will also be interesting to evaluate such a framework for deep underwater imaging as [2] observed that the thin refractive plane assumption well approximate for thick refractive interface.

Acknowledgements: This work was supported through an Innovative Engineering for Health award by The Wellcome Trust [WT101957], the Engineering and Physical Sciences Research Council (EPSRC) [NS/A000027/1, EP/N013220/1, EP/P012841/1] and the EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1).

References

[1] A. Agrawal and S. Ramalingam. Single image calibration of multi-axial imaging systems. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1399–1406, June 2013. 1

[2] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. A theory of multi-layer flat refractive geometry. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3346–3353, 2012. 1, 2, 3, 4, 8

[3] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*, 1966. 6

[4] Y.-J. Chang and T. Chen. Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In *IEEE Conference on Computer Vision*, pages 351–358, 2011. 2, 3

[5] V. Chari and P. Sturm. Multiple-View Geometry of the Refractive Plane. In *British Machine Vision Conference*, pages 1–11, 2009. 2, 4

[6] S. Chaudhury, T. Agarwal, and P. Maheshwari. Multiple view 3D reconstruction in water. In *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, pages 1–4, 2015. 3

[7] P. Daga, F. Chadebecq, D. Shakir, L. Garcia Peraza Herrera, M. Tella Amo, G. Dwyer, A. L. David, J. Deprest, D. Stoyanov, T. Vercauteren, and S. Ourselin. Real-time mosaicing of fetoscopic videos using sift. In *Proceedings of SPIE*, volume 9786. SPIE Medical Imaging, 2016. 8

[8] G. Dwyer, F. Chadebecq, M. Amo, C. Bergeles, E. Maneas, V. Pawar, E. V. Poorten, J. Deprest, S. Ourselin, P. De Coppi, T. Vercauteren, and D. Stoyanov. A continuum robot and control interface for surgical assist in fetoscopic interventions. *IEEE Robotics and Automation Letters*, 2(3):1656–1663, July 2017. 8

[9] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng. Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):930–943, Aug 2003. 6

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

[11] S. Haner and K. Aström. Absolute pose for cameras under flat refractive interfaces. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1428–1436, 2015. 3

[12] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 5

[13] A. Jordt, K. Köser, and R. Koch. Refractive 3d reconstruction on underwater images. *Methods in Oceanography*, 2016. 2, 3

[14] A. Jordt-Sedlazeck, D. Jung, and R. Koch. Refractive plane sweep for underwater images. In *Lecture Notes in Computer Science*, volume 8142, pages 333–342. 2013. 2, 3

[15] A. Jordt-Sedlazeck and R. Koch. Refractive Structure-from-Motion on Underwater Images. In *IEEE Conference on Computer Vision*, pages 57–64. 2013. 2, 3

[16] L. Kang, L. Wu, and Y.-H. Yang. Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model. *Applied optics*, 51(31):7591–7603, 2012. 1, 2

[17] L. Kang, L. Wu, and Y.-H. Yang. Two-view underwater structure and motion for cameras under flat refractive interfaces. In *European Conference on Computer Vision*, pages 303–316. 2012. 3

[18] J. Lavest, G. Rives, and J. Laprest. Underwater camera calibration. In *European Conference on Computer Vision*, volume 1843, pages 654–668. 2000. 1, 2, 5

[19] H. Li, R. Hartley, and J.-h. Kim. A linear approach to motion estimation using generalized camera models. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 5

[20] T. Łuczyński, M. Pfingsthorn, and A. Birk. The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings. *Ocean Engineering*, 133:9–22, 2017. 2

[21] M. Massot-Campos and G. Oliver-Codina. Optical sensors and methods for underwater 3d reconstruction. *Sensors*, 15(12):31525–31557, 2015. 2

[22] R. Pless. Using many cameras as one. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–587. IEEE, 2003. 2, 5

[23] S. Ramalingam, S. K. Lodha, and P. Sturm. A generic structure-from-motion framework. *Computer Vision and Image Understanding*, 103(3):218–228, 2006. 3

[24] A. Sedlazeck and R. Koch. Perspective and non-perspective camera models in underwater imaging, overview and error analysis. In *Lecture Notes in Computer Science*, volume 7474, pages 212–242. 2012. 1, 2, 3

[25] A. Shibata, H. Fujii, A. Yamashita, and H. Asama. Absolute scale structure from motion using a refractive plate. In *IEEE/SICE International Symposium on System Integration*, pages 540–545, 2015. 3

[26] H. Stewénius, M. Oskarsson, K. Aström, and D. Nistér. Solutions to minimal generalized relative pose problems. 2005. 5

[27] P. Sturm and J. Barreto. General imaging geometry for central catadioptric cameras. In *European Conference on Computer Vision*, pages 609–622. 2008. 1, 2, 4

[28] P. Sturm, S. Ramalingam, and S. Lodha. On Calibration, Structure-From-Motion and Multi-View Geometry for General Camera Models. In *International Society for Photogrammetry and Remote Sensing, Panoramic Photogrammetry Workshop*, volume XXXVI-5/W8, 2005. 5

[29] T. Treibitz, Y. Schechner, C. Kunz, and H. Singh. Flat Refractive Geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):51–65, 2012. 2

[30] S. You, R. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi. Waterdrop stereo. *arXiv preprint arXiv:1604.00730*, 2016. 1

[31] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, Oct 1994. 7