Refractive Two-View Reconstruction for Underwater 3D Vision

François Chadebecq[1] · Francisco Vasconcelos[1] · René Lacher[1] · Efthymios Maneas[1] · Adrien Desjardins[1] ·
Sébastien Ourselin[2] · Tom Vercauteren[2] · Danail Stoyanov[1]

Abstract

Recovering 3D geometry from cameras in underwater applications involves the Refractive Structure-from-Motion problem where the non-linear distortion of light induced by a change of medium density invalidates the single viewpoint assumption. The pinhole-plus-distortion camera projection model suffers from a systematic geometric bias since refractive distortion depends on object distance. This leads to inaccurate camera pose and 3D shape estimation. To account for refraction, it is possible to use the axial camera model or to explicitly consider one or multiple parallel refractive interfaces whose orientations and positions with respect to the camera can be calibrated. Although it has been demonstrated that the refractive camera model is well-suited for underwater imaging, Refractive Structure-from-Motion remains particularly difficult to use in practice when considering the seldom studied case of a camera with a flat refractive interface. Our method applies to the case of underwater imaging systems whose entrance lens is in direct contact with the external medium. By adopting the refractive camera model, we provide a succinct derivation and expression for the refractive fundamental matrix and use this as the basis for a novel two-view reconstruction method for underwater imaging. For validation we use synthetic data to show the numerical properties of our method and we provide results on real data to demonstrate its practical application within laboratory settings and for medical applications in fluid-immersed endoscopy. We demonstrate our approach outperforms classic two-view Structure-from-Motion method relying on the pinhole-plus-distortion camera model.

Keywords Underwater imaging · Two-view Refractive Structure-from-Motion · Flat refractive geometry

1 Introduction

Underwater imaging and 3D reconstruction are important for a variety of practical applications ranging from submarine exploration and robot guidance to fluid-immersed endo-scopic surgical procedures such as arthroscopy or fetoscopy (see examples shown in Fig. 1). The problem of recovering both the 3D geometry of the environment and the camera motion when operating under water requires adaptation of the normal pinhole-plus-distortion camera model because non-linear refractive distortion occurs at the interface between the liquid and the camera's housing. This means that the multi-view geometry of the problem is also influenced and additional considerations are required in order to formulate reconstruction as a Refractive Structure-from-Motion (RSfM).

When a camera is immersed in a fluid, light rays undergo refraction while passing through media with different optical density. The refractive distortion is defined by Snell's law and depends on the refractive index of traversed medium as well as the incidence angle of the incoming light ray (Hecht 1998). Dome windows can be used to efficiently compensate for refractive distortion. They however exhibit severe field curvature, they are unwieldy and need to be specially engineered (Luczyński et al. 2017). Flat-pane windows are

✉ François Chadebecq
f.chadebecq@ucl.ac.uk

1 Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), London, UK

2 School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
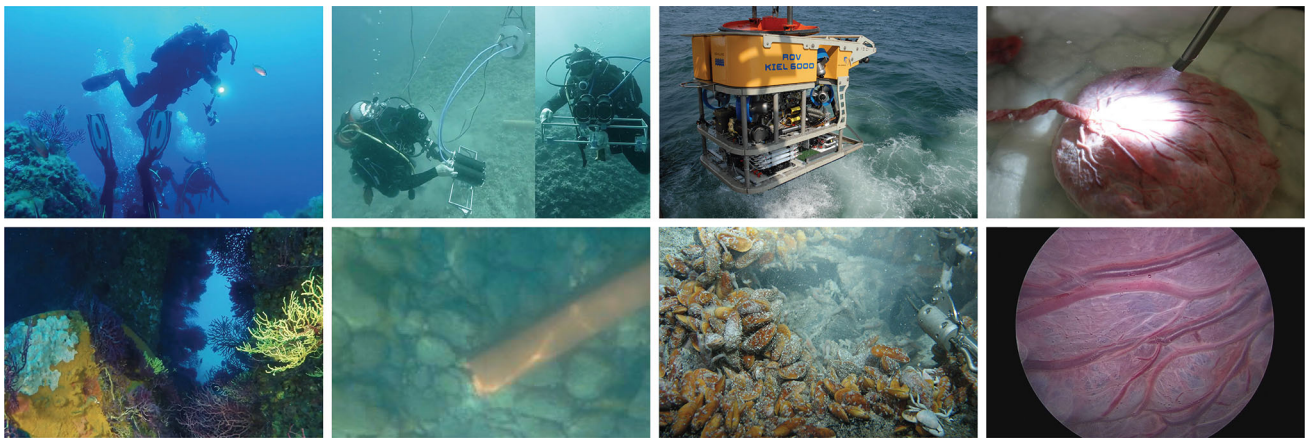
Fig. 1 Example of submarine and fluid-immersed computer vision applications. From the left column to the right: submarine exploration using a camera embedded in a waterproof casing (images courtesy of Mr. Giorgio Capelli, FiasParma, Federazione Italiana Attività Subacquee Parma, national park of Port-Cros and Porquerolles), underwater object recognition for the development of Autonomous Underwater Vehi-cles[images courtesy of Robotics and Intelligent Machines Laboratory, University of Parma, Marine Autonomous Robotics for InterventionS project (Lodi Rizzini et al. 2015; Oleari et al. 2015)], Remotely Operated Vehicle KIEL 6000 used for submarine exploration (images courtesy of GEOMAR Helmholtz Centre for Ocean Research Kiel), laboratory experiment reproducing fetoscopic examination of a human placenta

more flexible and affordable but they do not compensate for refractive distortion (see Fig. 2). Without taking into account refraction, it can be demonstrated that incoming light rays coming from the scene do not intersect anymore in a single center of projection and therefore the Single ViewPoint (SVP) assumption becomes invalid (Glaeser and Schröcker 2000; Treibitz et al. 2012, see Fig. 3a).

As a result, if a camera is calibrated in air assuming the pinhole model, calibration parameters cannot be used effectively once the camera is immersed within the liquid. While adapting the intrinsic camera parameters and distortion coefficients can compensate for refraction to some degree and for a limited working distance (Kang et al. 2012a; Treibitz et al. 2012), this approach introduces a systematic geometric bias which affects 3D measurements and camera pose estimation (Sedlazeck and Koch 2012). Indeed, refractive distortion directly depends on the incidence angle of the incoming light rays and as such on the scene depth [i.e. the 2D coordinates of a 3D point in the image plane directly depends on its depth (Treibitz et al. 2012)].

It has been demonstrated that image formation through a flat refractive interface can be modelled by an axial camera (Agrawal et al. 2012). However, such ray-based camera models are difficult to accurately calibrate and subsequently to use in practice due to the high dimensionality of their parametrisation (Glaeser and Schröcker 2000). The axial camera can also be regarded as a particular case of the generalised camera model for which multiple-view geometry has been extensively studied (Sturm et al. 2005; Sturm 2005; Li et al. 2008). However, this model is particularly sensitive to noise when considered for modelling monocular axial cameras whose image formation process can be efficiently approximated by a pinhole camera. Even if well calibrated, 3D reconstruc-

tion with a moving generalised camera remains even more challenging under water due to image quality degradations in such context. Approaches to solve this problem typically make prior assumptions, such as knowledge of the camera orientation (Chang and Chen 2011), or considering that the camera moves behind a fixed refractive interface (Chari and Sturm 2009) which is not suitable to estimate the motion of an immersed camera embedded within a waterproof housing.

An alternative approach to using generalised camera models is to explicitly consider one or more interfaces separating the optical system from the external medium. The media on either side of the interfaces have different refractive indexes causing refraction (Agrawal et al. 2012). By explicitly modelling this configuration's geometry, as shown in Fig. 3b, it is possible to form a refractive camera model with parameters that account for refractive distortion. Algorithms for 3D reconstruction can then rely on the refractive geometric constraints (Jordt-Sedlazeck et al. 2013; Jordt-Sedlazeck and Koch 2013; Jordt et al. 2016). Using refractive geometry, the two-view relative pose problem can be solved iteratively followed by bundle adjustment by associating each 2D image point to a virtual perspective camera (Agrawal et al. 2012). This formulation avoids directly computing the refractive reprojection error, which is computationally expensive as it requires solving a 12th degree polynomial (Treibitz et al. 2012). Dense 3D reconstruction can then be obtained by using a refractive plane sweep algorithm (Jordt-Sedlazeck et al. 2013).

More recently, Luczyński et al. (2017) proposed the Pinax camera model which combines a virtual pinhole camera model with the axial camera model. It more particularly allows the pre-computation of a lookup-table for fast and accurate refraction correction. The Pinax model has fur-

thermore demonstrated a high efficiency for stereo 3D reconstruction under water.

A notable current limitation of existing methods for estimating relative camera motion under water is that they are highly sensitive to noise and this results in a major practical limitation given that underwater images are subject to complex light scattering as well as medium turbidity, which makes precise point matches difficult to establish. In certain application domains, for example fluid-immersed endoscopic surgical procedures, accurate correspondence can be extremely challenging due to paucity of reliable features, medium properties and the restricted workspace of the moving camera. This means that stable and robust methods for estimating the refractive two view geometry and relative pose of cameras are an important area for further development in underwater vision and imaging.

Contribution

We propose a novel stereo 3D-reconstruction method for the case of a pinhole camera looking through a thin refractive interface, where the position and orientation of this interface are known (see Fig. 2). Our approach is particularly flexible and can be applied to various applications of underwater imaging and vision behind a thin refractive interface. This paper extends our previous work (Chadebecq et al. 2017) by:

i. providing with an explicit formulation of the refractive Fundamental relationship (rF), further developing the previous theoretical derivation of the two-view geometry for a fixed refractive plane (Chari and Sturm 2009) (see Sect. 4). By appropriately formulating the rF, we propose a robust stereo reconstruction method able to withstand complex underwater imaging conditions,

ii. extensively evaluating our approach for different underwater scenarios ranging from fluid-immersed medical endoscopy to deep underwater imaging. To demonstrate our contribution to the refractive geometry of underwater vision, we evaluate the effectiveness and improvements on numerical stability of our approach on synthetic and real data over classic two-view Structure-from-Motion (2View SfM) relying on the pinhole-plus-distortion perspective camera model. To highlight a particular application for two-view Refractive Structure-from-Motion (2View RSfM), we show results for fetoscopy, where an endoscope is used to guide vessel photo-coagulation on the placental surface within the amniotic fluid.

Appendices

For the sake of clarity, the refractive forward projection equation introduced in Agrawal et al. (2012) is recalled in
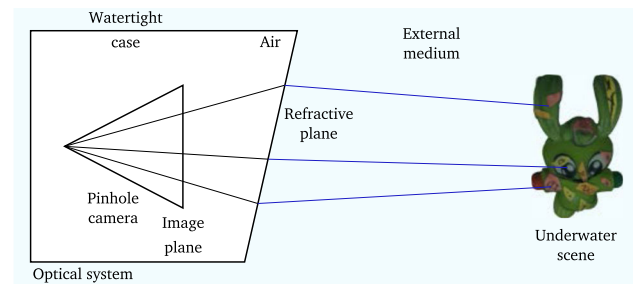

Fig. 2 Refractive camera model. We assume the underwater imaging system can be modelled by a pinhole camera embedded within a watertight case. The underwater scene is observed through a flat and thin refractive interface

appendix 1 and the refractive single-view geometry introduced in Chari and Sturm (2009) is recalled in "Appendix B".

2 Underwater Imaging Model and Refractive Structure-from-Motion: Related Work

To provide the necessary background to our work, we first introduce the camera models for refractive imaging which are generally considered for submarine and underwater exploration. A detailed review of these models can be found in Sedlazeck and Koch (2012). We then describe state of the art RSfM approaches. An exhaustive survey of these approaches can be found in Jordt (2014). A complementary survey on underwater 3D reconstruction can be found in Massot-Campos and Oliver-Codina (2015).

2.1 Refractive Image Formation Model

In previous works, three different image formation models have been used for refractive imaging and vision: the pinhole model (Kang et al. 2012a), the ray-based and generalised model (Chari and Sturm 2009), and the refractive model (Agrawal et al. 2012; Jordt 2014).

Adapted Pinhole Camera Model Most underwater computer vision literature has previously relied on the classic pinhole camera model of perspective projection. This choice is mainly motivated by its simplicity, general popularity and efficiency despite the known systematic geometric bias due to refraction (Glaeser and Schröcker 2000; Treibitz et al. 2012). Several papers have demonstrated that focal length and distortion parameters could be adapted to partially compensate for refraction effects (Fryer and Fraser 1986; Lavest et al. 2000; Kang et al. 2012a). Such assumption remains valid assuming the refractive plane is fronto parallel to the image plane of the camera. If it is tilted, symmetric or irregular distortion functions should be considered (You et al. 2016; Agrawal and Ramalingam 2013). Moreover, the adapted pinhole camera model is valid within a limited range of distance
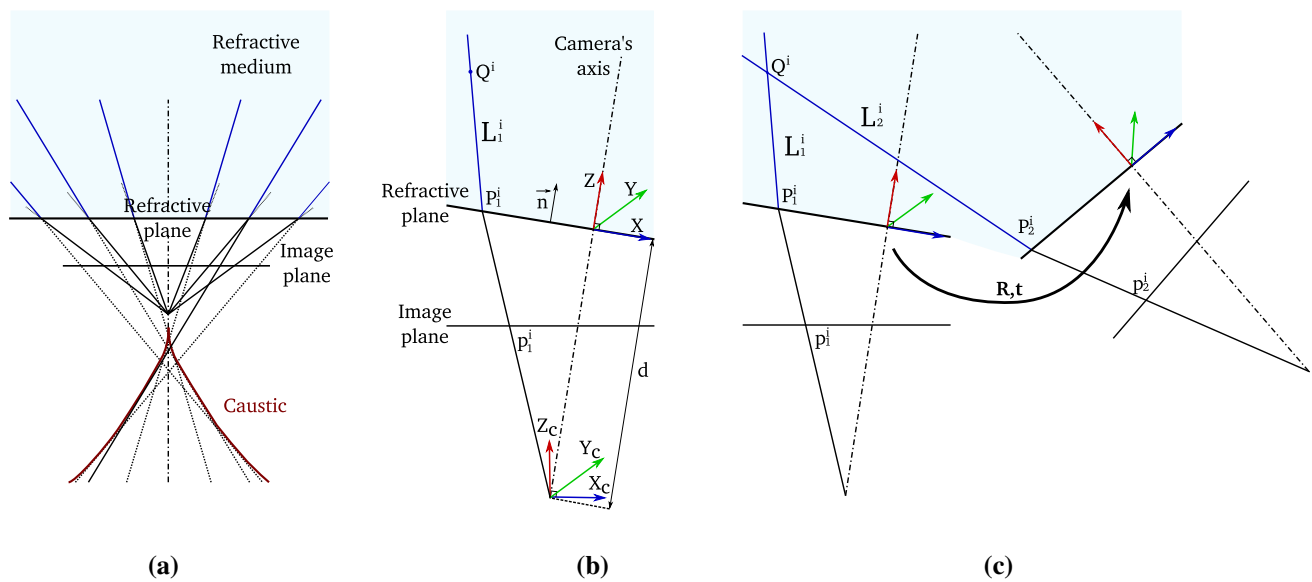
(a) **(b)** **(c)**

Fig. 3 **a** Illustration of the ray-based camera model, where the rays refracted at the refractive interface of the optical system form a caustic. Due to this refraction, the single viewpoint assumption is not valid. The axial camera model can be defined by the envelope tangent to the refracted rays. As such, the caustic is a measure of the deviation from the pinhole camera model. **b** Illustration of the refractive camera model assuming a single, thin refractive interface attached to the camera. We assumed the refractive plane position and orientation to be known. **c** Two-view refractive geometry assuming the refractive model from (**b**)

(the range of distances considered for camera calibration) as refractive distortion depends on the depth of the scene (Treibitz et al. 2012). The underlying geometry of rays actually defines an axial camera model (Agrawal et al. 2012). The influence of scene depth on the approximation error of the adapted pinhole model has been evaluated in Luczyński et al. (2017). They simulated an underwater scene observed at a reference distance of 2 m corresponding to the distance at which the camera was synthetically calibrated. They reported an approximation error of the order of 20 pixels for a 3D point located 50 cm away from the reference distance. This error corresponded to the average 2D distance between a set of 3D points projected within the image plane using the adapted pinhole and refractive camera models.

Ray-Based and Generalised Camera Model A complex but comprehensive approach is to associate each image pixel within the image to a light ray, similarly to the raxel model (Grossberg and Nayar 2005), and this has been used effectively in underwater imaging (Glaeser and Schröcker 2000; Sturm and Ramalingam 2004; Sturm et al. 2005; Treibitz et al. 2012). The imaging system can be characterized by its caustic, which is defined as the surface tangential to the bundle of rays intersecting the camera axis (see the illustration in Fig. 3a). In practice, this model is difficult to calibrate due to its large number of parameters (Sturm and Ramalingam 2004). An alternative is the generalised camera model which describes a generic and unified image formation process for both single and non-single viewpoint cameras (Grossberg

and Nayar 2001). Calibration of the generalised model and its use in multiple-view geometry have been studied (Sturm et al. 2005) and the approach has also been extended to underwater imaging assuming a camera moving behind a fixed refractive plane (Chari and Sturm 2009).

Refractive Camera Model Explicitly considering refraction at one or multiple flat-refractive planes (sharing the same orientation) has been proposed in a range of works (Agrawal et al. 2012; Jordt-Sedlazeck and Koch 2012; Yau et al. 2013; Chen and Yang 2014). We adopt this approach considering a single and thin refractive interface (illustrated in Fig. 3b). Such configuration is particularly adapted to underwater camera for which entrance lens is in direct contact with the external medium. This is for instance the case of fluid-immersed endoscopic imaging system used for medical procedure such as fetoscopy or arthroscopy. Moreover, it has been shown in Agrawal et al. (2012) that such relatively simple model (compared to the thick interface case) is capable of approximating both single and multiple-layer systems which allows us to propose a generic method for 2View RSfM.

Combined Camera Model The Pinax model proposed by Luczyński et al. (2017) combines a virtual pinhole camera model with the axial camera model. It assumes the distance between the refractive plane and the camera's optical center is close to the optimal distance for which the interval defined by the intersection of the light rays and the camera's axis is minimal. Under this assumption, it can be shown that the pinhole model efficiently approximates the underwater

image formation process. As such, each image point of the virtual pinhole camera can be back-projected onto a virtual pinax plane (virtual plane perpendicular to the camera's axis and situated at a reference distance in the scene) and then projected forward onto the physical camera (modelled by an axial camera). This allows the pre-computation of a lookup-table for fast and accurate refraction correction. Moreover, it is also particularly convenient as it only requires an in-air calibration of the underwater camera system. It is however sensitive to external medium refractive index changes and the distance of the scene can influence 3D reconstruction accuracy as it directly relies on the SVP approximation (which is an important limitation considering applications such as fluid-immersed endoscopy).

2.2 Refractive 3D Reconstruction

The problem of recovering 3D geometry in the presence of refraction has been studied in the past and there is a recent survey on active and passive 3D reconstruction methods for underwater imaging (Massot-Campos and Oliver-Codina 2015). In this section, we briefly review the most relevant approaches to this paper. The majority of approaches rely on standard SfM methods assuming the pinhole-plus-distortion camera model can compensate for refraction, which is known as the SVP assumption (Yu et al. 2003). It has been demonstrated that adapting radial lens distortion and focal length can compensate for refraction to a certain extent (Lavest et al. 2000) but this only leads to acceptable pose and 3D shape estimation within a limited range of distance of the scene (Kang et al. 2012b). There still exists a systematic geometric bias and estimated camera poses absorb some of the errors due to modelling (Sedlazeck and Koch 2012). Moreover, Jordt-Sedlazeck and Koch (2013) showed that measurement errors are particularly significant when the refractive interface is not strictly fronto-parallel to the image plane. As such, standard SfM methods remain poorly reliable under water due to drifting of camera poses estimations.

Few dedicated RSfM methods relying on ray-based models have been proposed. This is notably due to the difficulty of accurately calibrating such a system and its high sensitivity to noise. It has moreover been demonstrated that physical-based models outperform the generalised camera model (see Mouragnon et al. (2009) for SfM in the context of deep underwater imaging). Two theoretical works remain however particularly relevant to the refractive geometry of underwater vision. Considering the case of a camera moving behind a fixed refractive plane such as a camera seeing through an aquarium (fixed interface scenario), Chari and Sturm (2009) derived the rF relationship. It is however defined by a 12×12 matrix whose estimation is computationally unstable. An alternative ray-based model has been proposed in Chaudhury et al. (2015). It allows to express the underlying refractive

geometry as a direct extension of the projective geometry. The authors assume refraction happens at the camera centre only allowing them to geometrically estimate the relationship between the observed image point and the one obtained if refraction had not taken place [following a similar idea to the division model for distortion (Fitzgibbon 2001)]. By correcting image coordinates, the fundamental matrix can be directly estimated, while the external camera parameters and the 3D reconstruction are obtained up to a similarity.

Most of the RSfM approaches rely on a physical-based model by explicitly considering a pinhole camera looking through a refractive plane. The case of a fixed interface is considered in Chang and Chen (2011). They assume the camera embeds an Inertial Measurement Unit (IMU) and therefore both pitch and yaw of the camera are known. They proposed a closed-form solution to the absolute pose estimation, however it requires to know the vertical displacement of the camera. The case of two cameras embedded in different watertight casing is considered in Kang et al. (2012b). They propose an optimal solution to the relative translation problem under L_∞ assuming camera rotation is known. This solution is then extended to the unknown camera rotation case assuming a thin refractive plane parallel to both image plane of the cameras. More recently, Haner and Aström (2015) developed efficient minimal solvers for the cases considered in Chang and Chen (2011), Kang et al. (2012b) to solve for the absolute pose problem (camera observing known structure through a known refractive plane). For the general case of deep underwater imaging, Jordt et al. (2016) developed a complete RSfM framework by combining their previous works (Jordt-Sedlazeck and Koch 2013; Jordt-Sedlazeck et al. 2013). Relative camera motion between two successive views is estimated by relying on the flat refraction and coplanarity constraints derived from Agrawal et al. (2012) [respectively equations (12), (13)]. They then proposed a non-linear refractive bundle adjustment by extending (Ramalingam et al. 2006). It is based on the idea that a virtual camera can be associated to each 2D point into which the corresponding 3D point can be projected perspectively. Finally, dense depth estimation is obtained by using a refractive plane sweep algorithm (Gallup et al. 2007). This method requires a good initialisation, more particularly for two-view pose estimation, and relies on a costly iterative algorithm.

It is worth to note that RSfM has also been considered to solve for the absolute scale ambiguity inherent to SfM (Shibata et al. 2015a, b). Indeed, knowing the position and orientation of the interfaces theoretically yields to the absolute camera motion as relative pose is no more invariant to scale change in camera translation (Jordt-Sedlazeck and Koch 2013). The authors of Shibata et al. (2015a) suggest to place a thick refractive plate in front of a pinhole camera in order to infer the absolute scale of a scene. The coplanarity constraint is extended to two views and directly solved using

a least squares method. However, this approach is particularly sensitive to noise and requires at least 17 points correspondences. It has been experimentally observed in Jordt et al. (2016) that RSfM methods are too sensitive to noise to be able to infer the absolute scale of a scene. This is particularly the case considering a thin refractive interface as we noticed during our experiments.

3 Notation

For consistency, we adopt the mathematical notation used in Chari and Sturm (2009) (see Fig. 3b). The world coordinate frame (X, Y, Z) is arbitrarily set up in the first view. The Z-axis lies on the camera's axis defined as the line passing through the normal of the refractive interface ($\mathbf{n} = (0\ 0\ 1)^\top$) and the camera optical centre. It is important to note that we here distinguish between the camera's axis (i.e. the camera is composed of a pinhole camera looking through a refractive interface) and the pinhole camera's axis Z_c. The camera's and pinhole camera's axis are thus not necessarily aligned. The X and Y axes lie on the refractive plane and respectively align with the orthographic projection of X_c-axis and Y_c-axis of the camera coordinate frame onto the refractive plane. The pose of the camera in the first view is expressed as $\mathsf{P}_1 = \mathsf{R}_r^{-1}(\mathsf{I} - \mathbf{t}_r)$ where I is the 3×3 identity matrix, R_r is the 3×3 rotation matrix corresponding to the refractive plane

The lifted coordinates of the n-dimensional vector $\mathbf{v}$ can be obtained by vectorising the upper triangular part of the matrix $\mathbf{v}\mathbf{v}^\top$. Therefore, the expressions $\hat{\mathbf{v}} = (v_x{}^2\ v_x v_y\ v_y{}^2\ v_x v_z\ v_y v_z\ v_z{}^2)^\top$ and $\hat{\mathbf{v}} = (v_u{}^2\ v_u v_v\ v_v{}^2 v_u v_w\ v_v v_w\ v_w{}^2 v_u v_x\ v_v v_x\ v_w v_x\ v_x{}^2\ v_u v_y\ v_v v_y\ v_w v_y\ v_x v_y v_y{}^2\ v_u v_z\ v_v v_z\ v_w v_z\ v_x v_z\ v_y v_z\ v_z{}^2)^\top$ respectively define the lifted coordinates of the three-dimensional vector $\mathbf{v} = (v_x\ v_y\ v_z)^\top$ and the six-dimensional vector $\mathbf{v} = (v_u\ v_v\ v_w\ v_x\ v_y\ v_z)^\top$. If two vectors are related by a linear transformation T such as $\mathbf{v_2} = \mathsf{T}\mathbf{v_1}$, their lifted coordinates are related by the expression $\hat{\mathbf{v}}_2 = \mathsf{D}^{-1}\mathsf{S}(\mathsf{T} \otimes \mathsf{T})\mathsf{S}^\top \hat{\mathbf{v}}_1$ (Sturm and Barreto 2008). The symbol $\otimes$ refers to the Kronecker product. The design matrix S is a binary matrix of size $\frac{n^2+n}{2} \times n^2$. In the three-dimensional case, we denote S by S_t which is defined as:

$$S_\mathrm{t} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

In the six-dimensional case, we denote S by S_s. For the sake of brevity, we provide the matrix index corresponding to the non-null value (i.e. $\mathsf{S}_s{}^i_j$ corresponds to the entry in the ith row and jth column of the matrix).

$$\mathsf{S}_s \left(\begin{smallmatrix} 1 & 2 & 2 & 3 & 4 & 4 & 5 & 5 & 6 & 7 & 7 & 8 & 8 & 9 & 9 & 10 & 11 & 11 & 12 & 12 \\ 1 & 2 & 7 & 8 & 3 & 13 & 9 & 14 & 15 & 4 & 19 & 10 & 20 & 16 & 21 & 22 & 5 & 25 & 11 & 26 \end{smallmatrix} \right., \left. \begin{smallmatrix} 13 & 13 & 14 & 14 & 15 & 16 & 16 & 17 & 17 & 18 & 18 & 19 & 19 & 20 & 20 & 21 \\ 17 & 27 & 23 & 28 & 29 & 6 & 31 & 12 & 32 & 18 & 33 & 24 & 34 & 30 & 35 & 36 \end{smallmatrix} \right) = 1$$

orientation relative to the camera coordinate frame and $\mathbf{t}_r = (0\ 0\ d)^\top$ is the refractive plane position. The interface to camera centre distance along the camera's axis is denoted by d. A 2D image point i observed in view j is denoted by $\mathbf{p}^i_j = (x\ y\ 1)^\top$. We denote as $\mathbf{P}^i_j = (x\ y\ z\ 1)^\top$ the 3D point of incidence (point lying on the refractive interface) related to the 3D point $\mathbf{Q}^i$ projected in $\mathbf{p}^i_j$.

Light rays are defined by a starting point (e.g. a point of incidence) and a direction vector. Direction vectors associated to incident light rays (travelling within water) are denoted by $\mathcal{L}^i_j$. Direction vectors associated to their corresponding refracted light rays (travelling within the water tightness housing) are denoted by $\mathcal{P}^i_j$. They are directly derived, in the world coordinate system, from the expression $\mathcal{P}^i_j = (p_x{}^i_j\ p_y{}^i_j\ p_z{}^i_j)^\top = (\mathsf{R}_r^{-1} \frac{-\mathsf{K}^{-1}\mathbf{p}^i_j}{\|\mathsf{K}^{-1}\mathbf{p}^i_j\|})^\top$ where K^{-1} corresponds to the internal camera parameters of the pinhole camera. The Plücker coordinates of incident light rays are denoted by $\mathbf{L}^i_j = (\mathbf{L}_0{}^i_j, \ldots, \mathbf{L}_6{}^i_j)^\top$ (Yu et al. 2003). As such, $\mathbf{L}_{(a,b,c)}{}^i_j$ defines a vector composed of the elements $\mathbf{L}_a{}^i_j$, $\mathbf{L}_b{}^i_j$ and $\mathbf{L}_c{}^i_j$ of $\mathbf{L}^i_j$.

The matrix D is a diagonal matrix of size $m \times m$ where m is the number of rows of S. Its entries are given by the expression $\mathsf{D}_{ii} = \sum_{j=1}^k S_{ij}$ where k is the number of columns of S. Therefore, $\mathsf{D}_t = \mathrm{diag}(1\ 2\ 1\ 2\ 2\ 1)$ and $\mathsf{D}_s = \mathrm{diag}(1\ 2\ 1\ 2\ 2\ 1\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 1\ 2\ 2\ 2\ 2\ 2\ 1)$. These relationships can easily be extended to n dimensions.

4 Refractive Geometry

The refractive projection equation introduced in Chari and Sturm (2009) is recalled in "Appendix B". We here derive the two-view refractive relationship for the practical case of an underwater camera whose watertight case window can be modelled by a thin refractive plane.

4.1 Two-view Refractive Geometry

We here consider two views of an underwater scene. The incident light ray of direction-vector $\mathcal{L}^i_2$ stemming from the point $\mathbf{Q}^i$ is projected in $\mathbf{p}^i_2$ in the second view (see illus-

tration Fig. 3c). The Plücker coordinates of $\mathcal{L}_2^i$ are given by:

$$\mathbf{L}_2^i = \mathsf{T} \, (\lambda p_{x2}^i \ \lambda p_{y2}^i \ \gamma_{z2}^i \ -d\frac{p_{y2}^i}{p_{z2}^i}\gamma_{z2}^i \ d\frac{p_{x2}^i}{p_{z2}^i}\gamma_{z2}^i \ 0)^\top \qquad (1)$$

where $\mathsf{T} = \begin{pmatrix} \mathsf{R} & 0 \\ [t]_x \mathsf{R} & \mathsf{R} \end{pmatrix}$ and $[t]_x = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$ corresponds to the camera pose (Sturm and Barreto 2008). The scalar $\lambda = \frac{\mu_1}{\mu_2}$, where μ_1 corresponds to the refractive index within the camera housing (generally equal to 1) and μ_2 corresponds to the refractive index of the refractive medium. For brevity's sake we define $\gamma_{zj}^i = \sqrt{1 - \lambda^2 + \lambda^2 p_{zj}^{i\,2}}$.

By reordering the expression of $\mathbf{L}_2^i$, we can extract vector components $\mathbf{L}_{(6,1,2)2}^i$ and $\mathbf{L}_{(4,5,3)2}^i$:

$$\begin{pmatrix} \mathbf{L}_{(6,1,2)2}^i \\ \mathbf{L}_{(4,5,3)2}^i \end{pmatrix} = \mathsf{T}' \begin{pmatrix} \lambda t_u & 0 \\ 0 & t_v \end{pmatrix} \begin{pmatrix} \mathcal{P}_2^i \\ \gamma_{z2}^i \frac{\mathcal{P}_2^i}{p_{z2}^{i\,2}} \end{pmatrix} \qquad (2)$$

where $t_u = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $t_v = \begin{pmatrix} 0 & -d & 0 \\ d & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. The matrix T' corresponds to the camera pose matrix $\mathbf{T}$ reorganised:

$$\mathsf{T}' = \begin{pmatrix} R_{33} & [t]_{x3:}R_{.1} & [t]_{x3:}R_{.2} & R_{31} & R_{32} & [t]_{x3:}R_{.3} \\ 0 & R_{11} & R_{12} & 0 & 0 & R_{13} \\ 0 & R_{21} & R_{22} & 0 & 0 & R_{23} \\ R_{13} & [t]_{x1:}R_{.1} & [t]_{x1:}R_{.2} & R_{11} & R_{12} & [t]_{x1:}R_{.3} \\ R_{23} & [t]_{x2:}R_{.1} & [t]_{x2:}R_{.2} & R_{21} & R_{22} & [t]_{x2:}R_{.3} \\ 0 & R_{31} & R_{32} & 0 & 0 & 0 \end{pmatrix} \qquad (3)$$

The vector $R_{.i}$ identifies the ith column of the camera rotation matrix and we define $R_{.i} = [R_{1i} \ R_{2i} \ 0]^\top$. Similarly, $[t]_{xi:}$ identifies the ith row of the matrix $[t]_x$.

Expressing concisely the lifted coordinates of the aforementioned vector components is not obvious and eliminating the square root in γ_{z2}^i leads to a complex expression. We can nevertheless extract the lifted vector components $\left(\widehat{\mathbf{L}}_{(6,1,2)2}^{\top\,i} \ \widehat{\mathbf{L}}_{(4,5,3)2}^{\top\,i}\right)$ from the expression of $\left(\widetilde{\mathbf{L}_{(6,1,2)2}^{\top\,i} \ \mathbf{L}_{(4,5,3)2}^{\top\,i}}\right)$ obtained by vectorizing the upper triangular part of the matrix $\left(\mathbf{L}_{(6,1,2)2}^{\top\,i} \ \mathbf{L}_{(4,5,3)2}^{\top\,i}\right)$ $\left(\mathbf{L}_{(6,1,2)2}^{\top\,i} \ \mathbf{L}_{(4,5,3)2}^{\top\,i}\right)^\top$ (see Sect. 3). We can then derive the following relationship:

$$\left(\widehat{\mathbf{L}}_{(6,1,2)2}^{\top\,i} \ \widehat{\mathbf{L}}_{(4,5,3)2}^{\top\,i}\right) = \mathsf{N} \, \widehat{\mathsf{T}'} \, \widehat{\mathsf{C}} \left(\widetilde{\mathcal{P}_2^{i\,\top} \ \frac{\mathcal{P}_2^i}{p_{z2}^{i\,2}}}^\top\right) \qquad (4)$$

where N is a sparse change of basis binary matrix used to rearrange rows and reduce matrix dimensionality. It is defined by:

$$\mathsf{N}(\substack{1\,2\,3\,4\,5\,6\,7\,8\,9\,10\,11\,12 \\ 1\,2\,3\,4\,5\,6\,10\,14\,15\,19\,20\,21}) = 1 \qquad (5)$$

The lifted coordinates of the camera pose matrix T' are given by $\widehat{\mathsf{T}'} = \mathsf{D}_s^{-1} \mathsf{S}_s \mathsf{T}' \otimes \mathsf{T}' \mathsf{S}_s^\top$. Matrix D_s and S_s are defined in Sect. 3.

The matrix C, which notably embeds the refractive parameters (interface position and medium refractive index), is defined by:

$$\mathsf{C} = \begin{pmatrix} \lambda \mathsf{t}_u & 0 \\ 0 & \gamma_{z2}^i \mathsf{t}_v \end{pmatrix} \qquad (6)$$

The lifted matrix $\widehat{\mathsf{C}}$ is derived using $\widehat{\mathsf{C}} = \mathsf{D}_s^{-1} \mathsf{S}_s \mathsf{C} \otimes \mathsf{C} \mathsf{S}_s^\top$.

Combining equations (22) and (4) allows us to define the refractive fundamental relationship:

$$\left(\widetilde{\mathcal{P}_2^i \ \frac{\mathcal{P}_2^i}{p_{z2}^i}}\right)^\top \mathsf{rF} \left(\begin{array}{c} \widehat{\mathcal{P}_1^i} \\ \frac{\widehat{\mathcal{P}_1^i}}{p_{z1}^{i\,2}} \end{array}\right) = 0 \qquad (7)$$

The matrix rF is defined as:

$$\mathsf{rF} = \widehat{\mathsf{C}}^\top \widehat{\mathsf{T}'}^\top \mathsf{N}^\top \mathsf{M} \mathsf{D}_s \qquad (8)$$

where the matrix M is defined by:

$$\mathsf{M} = \begin{pmatrix} (1-\lambda^2)\mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_b^\top \otimes \mathsf{t}_b^\top \mathsf{S}_t^\top & 0 \\ \lambda^2 \mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_b^\top \otimes \mathsf{t}_b^\top \mathsf{S}_t^\top & -\lambda^2 \mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_a^\top \otimes \mathsf{t}_a^\top \mathsf{S}_t^\top \end{pmatrix} \qquad (9)$$

The matrix $t_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $t_b = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -d & 0 \\ d & 0 & 0 \end{pmatrix}$. The complete derivation of the matrix M is recalled in "Appendix B" (Chari and Sturm 2009).

The rF matrix is of size 12×21. It directly depends on refractive interface parameters (distance and orientation), as well as, the camera's motion.

5 Refractive Camera Pose Estimation

Directly estimating rF matrix entries from two-view correspondences is computationally unstable and practically difficult because it requires matching $12 \times 21 = 252$ points across the image pair. The high degree of freedom of the rF matrix prevents closed-form algebraic solutions to provide

with reliable estimates of its entries. These solutions are particularly sensitive to noise, as it can also be observed for the estimation of the generalised essential matrix considering a monocular axial camera. Even using a robust method such as RANdom Sample Consensus (RANSAC), we were not able to reliably estimate the rF matrix and as such extract absolute or relative cameras' poses.

This suggested to rely on a two step relative pose estimation method. The relative camera motion is first estimated using the classic two-view SfM pipeline. Assuming the refractive geometry of the camera is calibrated (using Agrawal and Ramalingam 2013), its pose can then be refined using both the rF relationship and reprojection constraints. The rF matrix is computed thanks to equation (8). The reprojection constraint is based on the refractive forward projection equation (see "Appendix A"). It is worth to note that the rF constraint is also essential to remove outlier correspondences. This is a particularly sensitive key point in underwater imaging where images are highly affected by medium properties.

5.1 Camera Calibration

We assume the camera to be calibrated both in air and under water. In air, calibration has been achieved using (Bouguet 2008). Underwater calibration consists in calibrating the camera considering both the single viewpoint approximation and the refractive camera model. The underwater pinhole calibration is similar to the calibration in air. Calibration of the position and orientation of the thin refractive plane has been performed using (Agrawal et al. 2012) relying on the same set of images. Whereas this method only requires a single view of the calibration target, we relied on the full set of underwater calibration images (20 images). We experimentally noticed a stable estimation of the interface orientation while the estimation of the interface depth suffered from a significant variance. We used the average interface depth estimated over the 20 underwater calibration images.

It is worth to note that calibrations have been performed under water before each experiment. As it will be showed using synthetic datasets in Sect. 6.1, slight changes in medium refractive index do not significantly impact camera pose estimation.

5.2 Underwater Relative Pose Estimation

The proposed relative pose estimation method is depicted in Fig. 4.

Our method is composed of the following steps. The first step relies on the single viewpoint assumption. After undistorting images using distortion parameters estimated under water, SIFT features are extracted and matched between each pair of views (Lowe 1999; Vedaldi and Fulkerson 2008; Daga
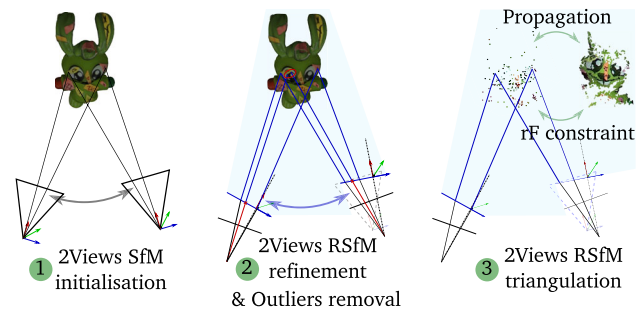

Fig. 4 Overview of the proposed relative pose estimation method. The relative pose is first estimated using the classic fundamental constraint assuming the camera has been calibrated under water. The relative camera pose is then refined using the refractive Fundamental (rF) and reprojection constraints. Outlier Correspondences are discarded based on the rF constraint. Propagating key correspondences using (Stoyanov et al. 2010) and removing geometrically inconsistent correspondences using the rF constraint allowed us to obtain quasi dense reconstruction

et al. 2016). Along this step, fundamental matrix and relative camera motion are estimated using standard direct estimation approaches (Yu et al. 2003). The second step consists in refining the single viewpoint approximation. This is done by minimising the following non-linear least squares constraint based on the rF and reprojection constraints:

$$\theta^*, \mathbf{t}^* = \arg\min_{\theta, \mathbf{t}} \sum_{i=1}^{n} FC_i(\theta, \mathbf{t}) + \sum_{i=1}^{n} Proj_i(\theta, \mathbf{t}) \qquad (10)$$

where the Euler angles $\theta = (\theta_x, \theta_y, \theta_z)^\top$ and $\mathbf{t}$ define the rotation and translation of the camera respectively. The indices $i = \{1, \ldots n\}$ correspond to the point number. FC_i corresponds to the refractive fundamental constraint defined in equation (7) and computed for the point i. The reprojection constraint $Proj_i$, computed for the point i, is defined by:

$$Proj_i(\theta, \mathbf{t}) = \|\mathcal{P}_p \mathcal{P}_r(\theta_0, \mathbf{t}_0, \mathbf{Q}^i) - \mathbf{q}_1^i\|_2^2 +$$
$$\left\|\mathcal{P}_p \mathcal{P}_r(\theta, \mathbf{t}, \mathbf{Q}^i) - \mathbf{q}_2^i\right\|_2^2 \qquad (11)$$

where $\mathbf{q}_1^i$ and $\mathbf{q}_2^i$ respectively correspond to the observations of the 3D point $\mathbf{Q}^i$ in the first and second view (i.e. corresponding to matching i). The refractive projection funtion $\mathcal{P}_r$ projects the 3D point $\mathbf{Q}^i$ onto the refractive plane. The refractive forward projection function is derived in Agrawal et al. (2012). Its definition is recalled in appendix 1, equation (21). We denote as $(\theta_0, \mathbf{t}_0)$ the camera's pose corresponding to the first view. Without any loss of generality, we fixed $\theta_0 = \mathbf{t}_0 = (0, 0, 0)^\top$. Points lying on the refrative plane are projected onto the image plane thanks to the perspective projection matrix $\mathcal{P}_p$ (obtained by calibration).

We used a similar weight for the reprojection and the refractive fundamental constraints involved in cost function (10) as we experimentally observed that it allowed us to obtain the most accurate pose and 3D reconstruction estimates. As demonstrated by our experiments, the 2View SfM pose estimates computed in the first step of our method provided us with an accurate and reliable initialisation of camera poses sufficiently close to the optimum estimates. The minimisation problem is then solved using the Levenberg-Marquardt method (Marquardt 1963) and the rF matrix is used to reject outliers. We practically used a threshold of 0.1 pixels to truncate matches that do not conform to the refractive geometry.

We finally obtain quasi-dense stereo correspondences using an iterative growing scheme (Stoyanov et al. 2010). Again, outliers are discarded using the refractive geometry which leads to a final quasi-dense reconstruction.

5.3 Discussion

The proposed 2View RSfM approach explicitly relies on refractive geometry. Unlike 2View SfM, relying on the adapted pinhole model, it does not suffer from a systematic geometric bias. Additionally, it allows us to efficiently discard mismatches which is of particular significance for application such as fluid-immersed endoscopy for which underwater artefacts generally prevent the reliable tracking of features (complex light scattering, medium turbidity, motion blur due to the manipulation of a flexible endoscope within a closed environment). By combining our approach with the method proposed in Stoyanov et al. (2010), we obtain quasi-dense reconstructions.

The proposed method nevertheless suffers from limitations. Because its second step relies on a costly non-linear minimisation (equation (10)), it is not real time. Furthermore, as its first step relies on 2View SfM it inherits from its limitations. Most particularly, it is sensitive to the same critical camera motions and it suffers of scale ambiguity (Yu et al. 2003). Similarly, the 3D densification step inherits from Stoyanov et al. (2010) method drawback. As such, it is sensitive to specular reflections and model occlusion.

6 Experiments

The proposed 2View RSfM method has been evaluated on both synthetic and real data. Synthetic evaluation considers general underwater scenario where the scene is generally situated at a distance of 2 to 4 meters to the camera. For real experiments, we consider the scenario of a consumer action camera imaging a scene situated at a close distance of approximately 0.5 meters. We also highlight a particular application for 2View RSfM and show results for fetoscopy. In such con-
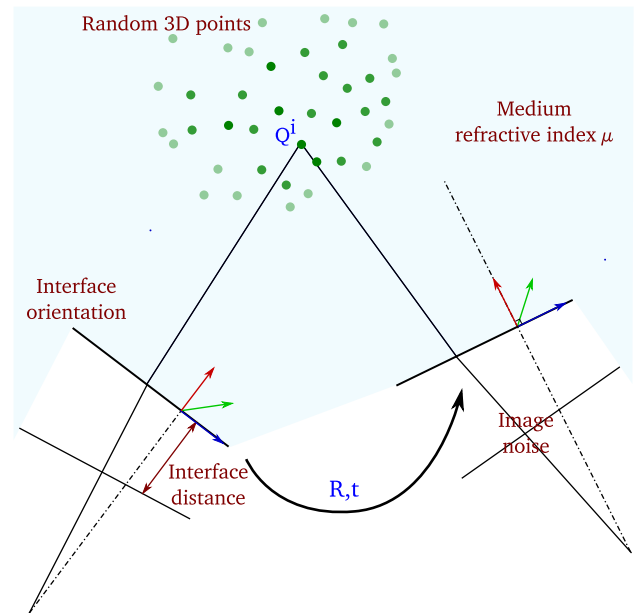

Fig. 5 Synthetic evaluation. A set of randomly generated 3D points is projected into two views using the refractive forward projection equation (Agrawal et al. 2012). Camera motion is randomly generated. We evaluate the influence of medium refractive index, interface position and orientation as well as image noise on the robustness of our method. Parameters written in red correspond to known parameters or parameters for which we want to estimate the influence on two-view Refractive Structure-from-Motion. Parameters written in blue correspond to unknown parameters

text, providing the surgeon with 3D reconstruction (as well as endoscope displacement) could facilitate both diagnosis and intervention within highly sensitive cavity such as uterus.

Our code and used datasets will be published online at http://www.surgicalvision.cs.ucl.ac.uk/resources/code/.

6.1 Synthetic Evaluation

The synthetic dataset has been generated using the following steps (see Fig. 5). A set of 3D points situated at a distance between 2 and 4 meters from the camera are randomly generated. They are assumed to belong to a smooth 3D surface of approximate dimensions 1×1 meter. For each setup, we generated 100 pairs of synthetic images considering a virtual camera moving in front of the scene. We used the refractive forward projection equation (21) derived in Agrawal et al. (2012). We considered the following internal camera parameters; focal length: 800 pixels, principal point: 640×480 pixels (no distortion parameters, no skew, image size: 1280×960 pixels). The refractive parameters have been randomly chosen for each pair of view. The refractive plane orientation lays in the interval $-\pi/6$ to $\pi/6$ radians (refractive plane normal rotation) and its depth in the interval 2 to 15 millimetres (mm). The camera rotation lays in the interval $-\pi/6$ to
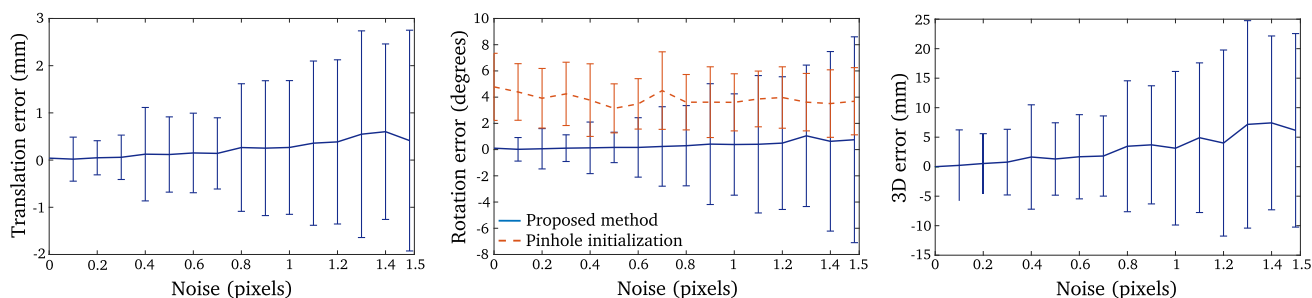
Fig. 6 Evaluation of estimation of camera motion in the presence of noise: camera translation estimation error (left), camera rotation estimation error (center), 3D reconstruction error (right)

$\pi/6$ radian and its translation in the interval 0.2 to 0.5 meters (absolute distance).

A set of 40 calibration images has also been synthetically generated. We assumed a synthetic checkerboard situated at a distance of 2 to 4 meters from the camera. We generated 20 synthetic calibration images of the checkerboard in air assuming the pinhole camera model whose internal parameters have been described in the previous paragraph. We added 0.2 pixels Gaussian noise to the synthetic images of the checkerboard simulating a real calibration scenario. We also generated 20 calibration images under water (using the method described for generating the images pairs dataset) and proceed to underwater calibration considering the single viewpoint assumption (also adding a Gaussian noise of 0.2 pixels). Refractive plane calibration has not been directly carried out, as we evaluated the influence of refractive plane calibration in our experiments, it has been synthetically controlled.

We evaluated the influence of noise, medium refractive index changes and refractive plane calibration accuracy on the robustness of camera motion estimation and accuracy of 3D reconstruction (see Fig. 5). We chose to present the median and median absolute deviation errors as we noticed cases of inconsistent pose estimation using the single viewpoint assumption for noise level greater than 0.5 pixels. Considering a noise level of 1.5 pixels it appeared in 12% of cases. Although the proposed method allowed us to withdraw outliers, results take into account the entire set of points.

Robustness to Noise we evaluated the robustness of our method to image noise. We considered a random Gaussian noise whose standard deviation lays within the range 0 to 1.5 pixels. Results are presented in Fig. 6. Despite significant noise levels, our method leads to accurate estimations of camera motions. The rotation error is less than 2 degrees despite a noise of 1.5 pixels and the translation error is approximately of 0.5 mm for the same level of noise. However, we observed a significant median absolute deviation error for noise level greater than 0.8 pixels. This was due to inconsistent initialisation leading to degenerate pose estimation. Considering a noise of 0.4 pixels we noticed 4 cases for which an inaccurate
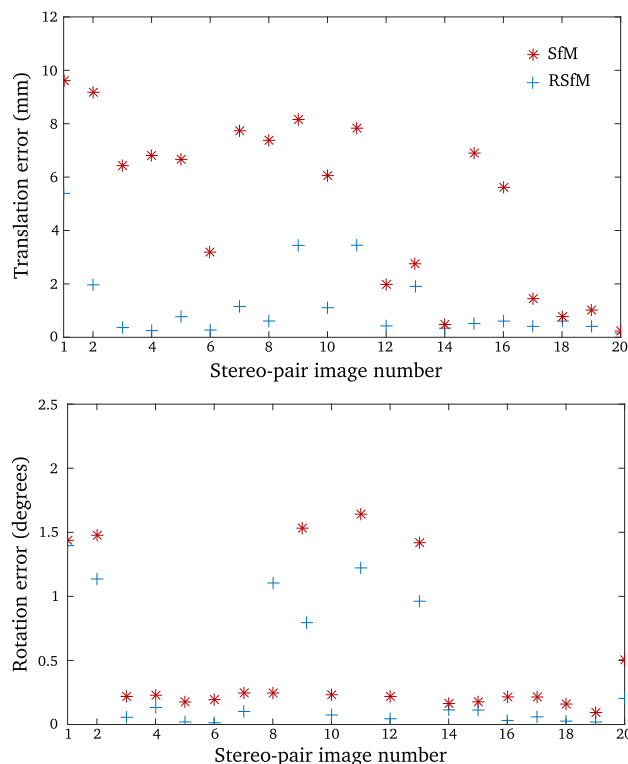

Fig. 7 Relative pose estimation between the cameras pair of a surgical stereo-endoscope. Despite the small baseline between the two cameras, two-view Structure-from-Motion pose estimations are efficiently refined considering the refractive camera model. This is here particularly noticeable for translation estimation

initialisation led to an alteration of camera pose estimation. For a noise between 1 and 1.5 pixels, such initialisation failures appeared in 10% of evaluated cases.

Despite accurate camera motions estimations, it is important to note that above 0.6 pixels of noise, the 3D reconstruction error is greater than 2 mm. Such error can be of significance considering medical applications. This is however related to the scale of the scene as well as the setup considered in our synthetic evaluation. The evaluation of the accuracy of 3D reconstruction in the context of fluid-immersed endoscopy is highlighted in the next section.
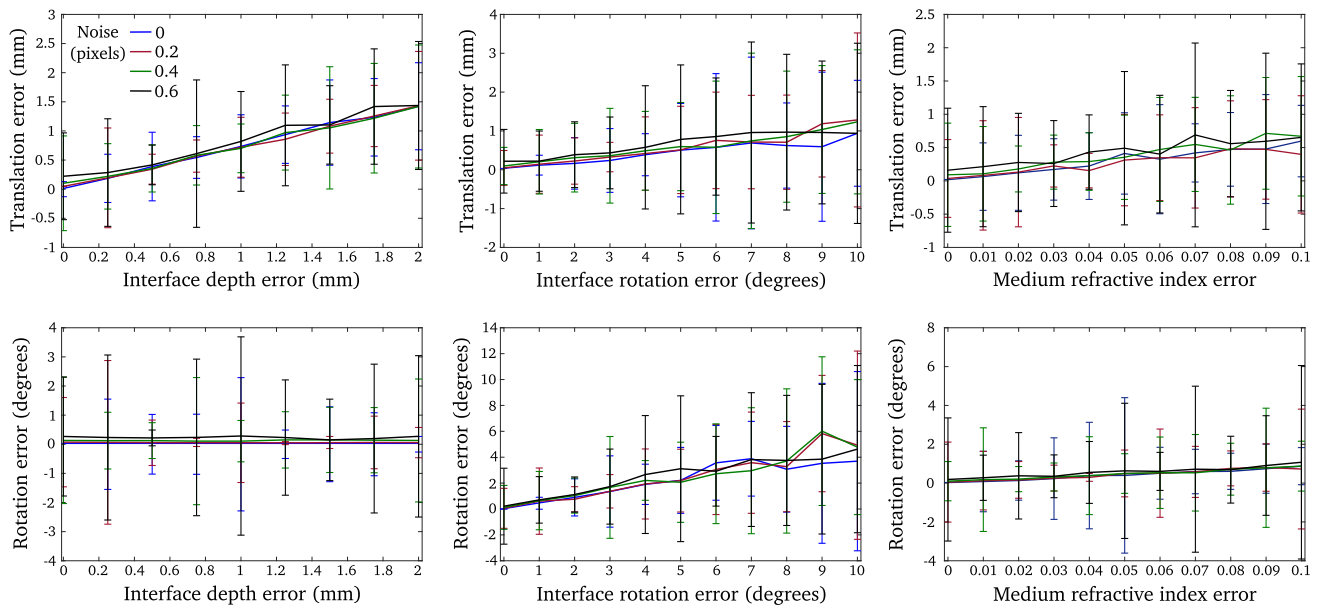
Fig. 8 Evaluation of camera motion estimation for inaccurate calibration of the refractive interface depth (left column), the refractive interface orientation (centre column) and the refractive medium (right column). The top line presents camera translation estimation error while the second line presents camera rotation estimation error

Robustness to inaccurate interface calibration we first highlight the influence of inaccurate interface depth calibration on relative pose estimation (see Fig. 8, left column). The interface depth error is evaluated in the interval 0 to 2 mm. We observed that an interface depth error greater than 1 mm has a significant impact on translation estimation. The translation error is greater than 1 mm for an interface depth calibration error greater than 1.4 mm. Such errors imply inaccurate 3D reconstructions and measurements as it has been previously outlined (see Fig. 6). Interface depth errors have a minimal impact on camera rotation estimations. Moreover, we observed that noise level lower than 0.6 pixels does not significantly impact on results and behaviour of our method.

We then highlight the influence of inaccurate interface orientation calibration (see Fig. 8, central column). Refractive interface orientation accuracy has a strong influence on camera pose estimation. The translation error is greater than 0.5 mm when the interface rotation error is greater than 5 degrees. The error of camera rotation estimation is approximately 2 degrees and 4 degrees for an interface rotation error greater than 9 degrees.

Robustness to medium refractive index changes this evaluation is particularly relevant in the context of fluid-immersed endoscopy. For instance, most computer vision approaches for medical imaging assume organic fluid refractive index is similar to the refractive index of water while the influence of such assumption has not been quantitatively evaluated. It is particularly important for procedure such as fetoscopy which consists in inserting an endoscope (namely fetoscope) within the uterine cavity during pregnancy in order to allow access to the fetus, the placenta and the amniotic cavity. The refractive index of amniotic fluid can evolve according to gestation time despite it is mainly composed of saline fluid (Steigman et al. 2010). However, estimating the refractive index of amniotic fluid during the procedure is complex. We considered the medium refractive index error lays between 0 and 0.1. We noticed that, even for important changes of fluid refractive index, the translation estimation error is less than 0.5 mm and the rotation error less than 1 degree. These results remain valid provided that the image noise level is lower than 0.6 pixels. We therefore observed that approximating organic fluid refractive index by water refractive index remains a fair assumption despite significant changes of medium properties.

6.2 Real Data Evaluation

6.2.1 Camera Pose Estimation Evaluation

For evaluating relative camera pose estimation, we carried out real experiments considering a surgical stereo-endoscope. We first calibrate the stereo-endoscope using a a set of 20 calibration images acquired outside water. We used a planar checkerboard and estimated both internal camera parameters and rigid pose between the cameras pair (Zhang 2000). We then followed the same procedure to calibrate the stereo-endoscope under water considering as such the adapted pinhole camera model (Nikitichev et al. 2017). We
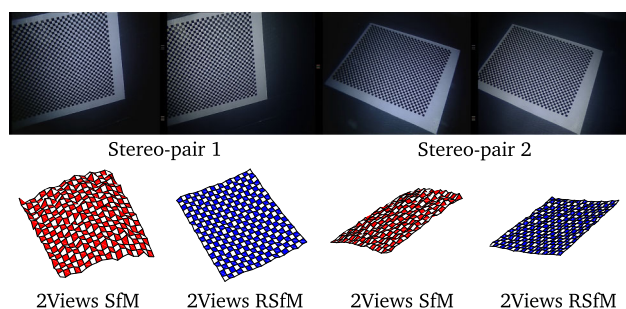
Fig. 9 Underwater 3D reconstruction of a planar checkerboard pattern using the two views of a surgical stereo endoscope. The proposed two-view Refractive Structure-from-Motion significantly improves the initial 3D shape estimate provided by two-view Structure-from-Motion

Fig. 10 Setup used to acquire the statuette and toy dataset. The statuette dataset was realised using a Go pro hero 3 camera embedded within its dedicated watertight casing. The toy dataset was realised using a Storz® Hopkins 26008 BA fetoscope

used the same set of underwater images to calibrate the orientation and distance of the flat refractive plane considering the flat refractive camera model (Agrawal et al. 2012). We finally collected a set of 20 additional images under water and compared relative camera pose estimation between the two images of the stereo pair using 2View SfM and 2View RSfM. We considered as the ground truth the rigid pose estimated during the first calibration step realised outside water. The translation between the two cameras of the stereo pair is 5.8 mm and the rotation is 3.5 degrees along the Y axis and less than a degree along the X and Z axis.

Results are reported in Fig.e 7. We observed a mean rotation error of 0.54 degrees for 2View SfM with a variance of 0.58 degrees. Using 2View RSfM slightly improved rotation estimation, we reported a mean rotation error of 0.35 degrees and a variance of 0.51 degrees. We noticed a significant improvement of translation estimation which strengthens the results obtained for the synthetic experiments. The mean translation error for 2View SfM is 5.24 mm with a variance of 3.21 mm. The mean translation error for 2View RSfM is 1.21 mm with a variance of 1.39 mm. The translational error observed for 2View SfM can be explained by the small baseline between the cameras. However, the initial relative pose estimation provided by 2View SfM can be efficiently refined using 2View RSfM. As can be seen in Fig. 9, it has a noticeable impact on 3D shape estimation.

6.2.2 3D Shape Estimation Evaluation

We evaluated our approach on three datasets. For each pair of views in our datasets, we used the same feature points for both 2View SfM and 2View RSfM. Outliers have been withdrawn thanks to the method described in section 5.2. The first dataset consists of 10 images of a statuette immersed within a tank filled with water (see Fig. 10). We used a Gopro® hero 3 camera embedded within its dedicated water tight casing. The size of the statuette is approximately 150 mm
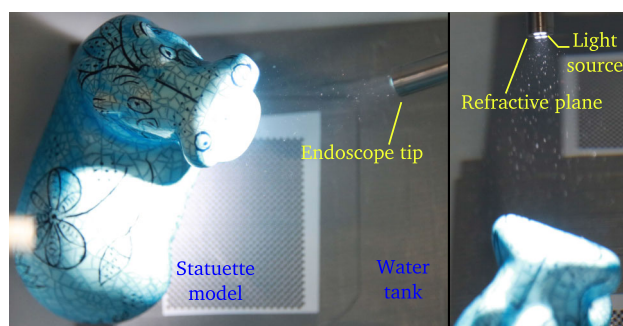
width and 140 mm height and the camera was situated at a distance of approximately 400 to 500 mm. The ground truth 3D object model was obtained using an Artec® Spider 3D scanner. The second dataset consists of 15 images of a toy immersed under identical conditions. Images were acquired using a Storz® Hopkins 26008 BA fetoscope. This endoscope model relies on a rod lens optical system. Therefore, it allowed us to validate that the single refractive plane assumption approximates well a complex optical system. The size of the toy is approximately 18 mm width and 15 mm height (figurine head). The distance from the endoscope distal end to the object was approximately 20 to 50 mm. Similarly to the statuette dataset, the ground truth 3D object model was obtained using an Artec® Spider 3D scanner. The third dataset consists of 10 images of a term human placenta collected following a caesarean section delivery and immersed under water (a written informed consent was obtained at University College London Hospital (UCLH) after the Joint UCL/UCLH Committees on the Ethics of Human Research approved the study (14/LO/0863)). Images were acquired using the fetoscopic equipment previously mentioned (distance between 50 to 80 mm). Our goal was to evaluate the robustness of our approach towards 2View SfM in real fetoscopic conditions. Fetoscopy is performed during pregnancy. The placenta positions itself at either the top or side of the uterus within the amniotic sac. At the early stage of pregnancy, the amniotic sac is filled with amniotic fluid which is progressively replaced by fetal urine. As such, the placenta is always immersed within fluid during pregnancy but the nature of this fluid evolves. Unfortunately, it was impossible to obtain a reliable ground truth 3D model of the placenta. Indeed, once the placenta is immersed within water, refractive distortion prohibits the use of 3D laser scanners. Furthermore, water penetrates within the organ significantly deforming it which invalidates any 3D model computed outside water.

Statuette dataset Figure 11 depicts a representative sample of the 3D reconstruction results obtained with both
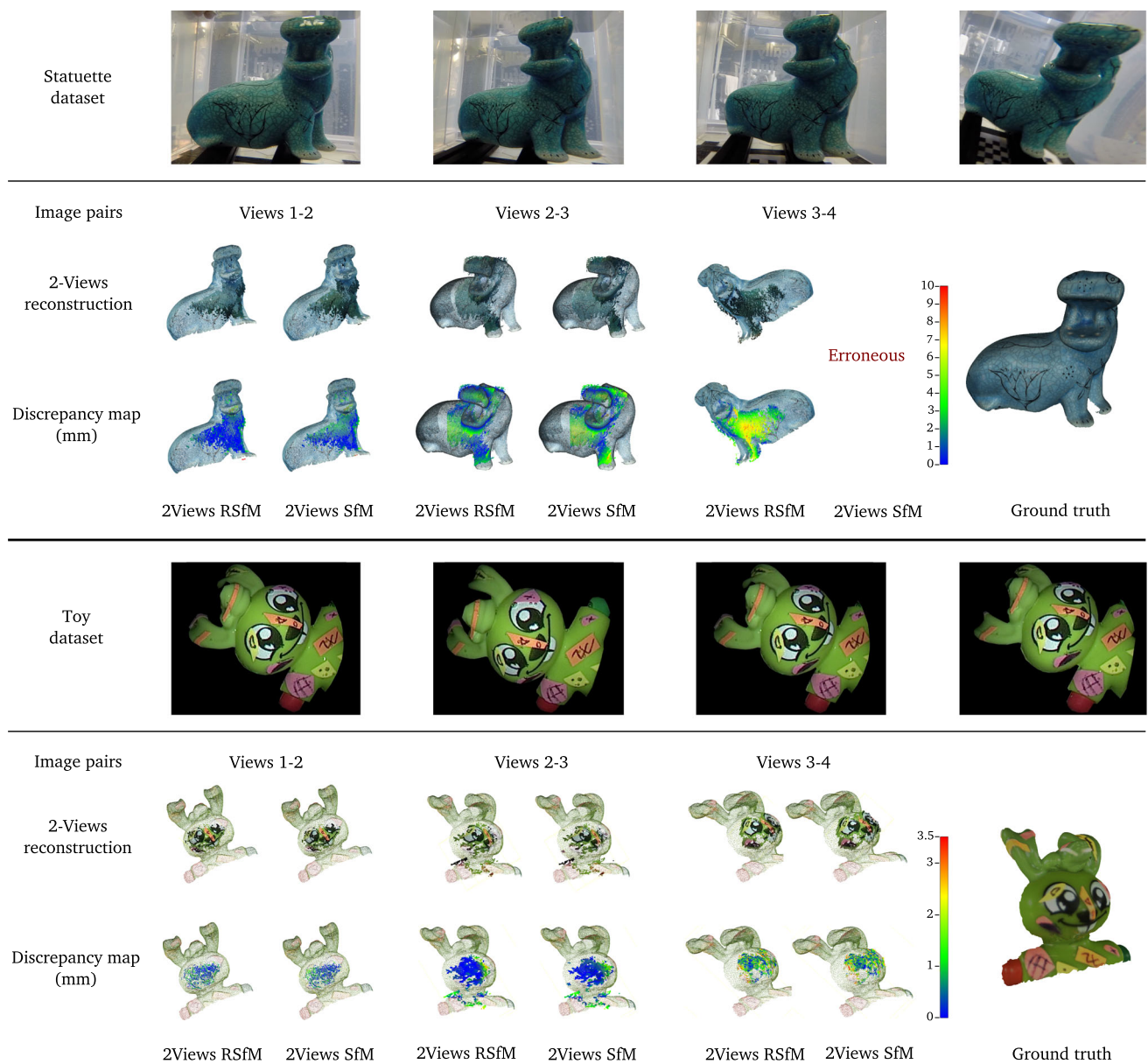
Fig. 11 Representative sample of the statuette and toy dataset results. The top row presents three consecutive images pairs as well as the ground truth 3D reconstruction obtained thanks to a high definition Artec® Spider 3D scanner. For each of the image pair we present the results obtained with the proposed two-view Refractive Structure-from-Motion method as well as two-view Structure-from-Motion method. The second row presents 3D reconstruction results overlay on the ground truth mesh. The last row presents the discrepancy map obtained after aligning 3D reconstruction results with the ground truth mesh. Results demonstrate a sensitive improvement of 3D reconstruction accuracy using two-view Refractive Structure-from-Motion. It is worth to note for the statuette dataset that, for two cases, two-view Structure-from-Motion did not allow us to obtain reliable 3D shape reconstruction

2View SfM (assuming underwater camera parameters) and our 2View RSfM method. More particularly, we present the discrepancy map between the ground truth reconstruction and these reconstructions. We observed a mean error of 5.3 mm and a standard deviation of 3.6 mm for 2View SfM while we obtained an error of 4.6 mm and a standard deviation of 3.2 mm for 2View RSfM. Beyond the accuracy improvement we noticed two particular cases where 2View SfM did not allow us to obtain reliable 3D reconstructions. Initial camera pose estimates were too erroneous in those cases to allow us to infer acceptable 3D shape of the statuette. We noticed that it corresponded to cases for which the object depth range was particularly large (side view of the statuette) or cases for which the object was located at a close distance from the camera. We nevertheless observed in such situations that our
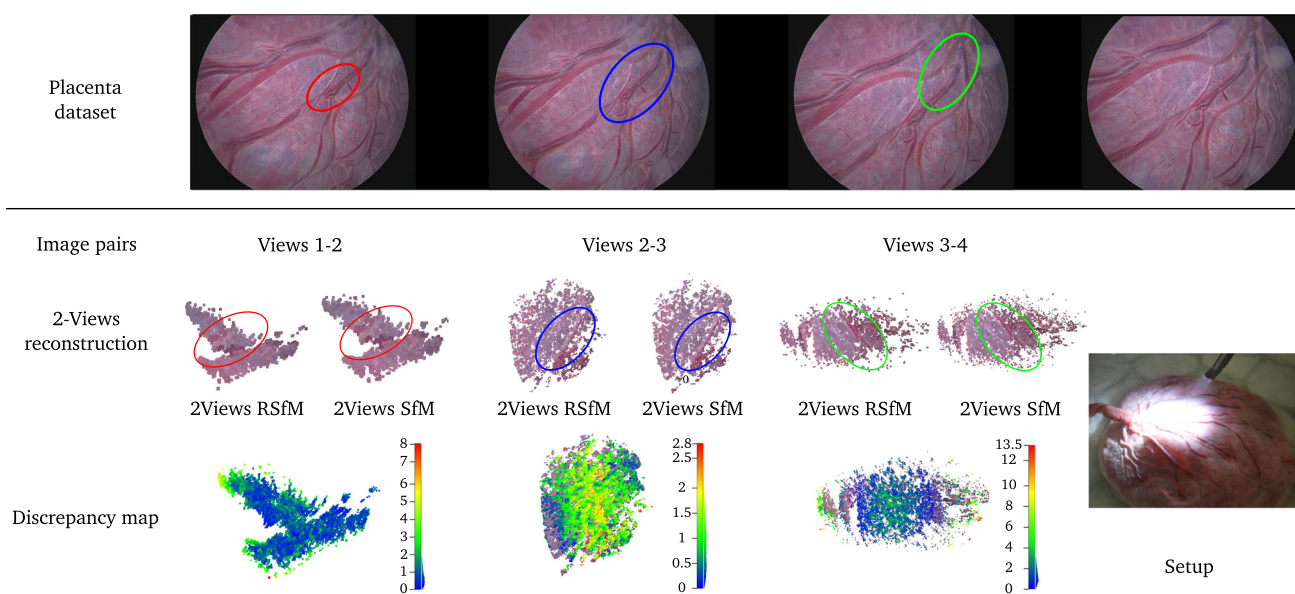
Fig. 12 Underwater 3D reconstruction of a placenta. The top row presents three consecutive images pairs of the placenta dataset. It has been obtain by imaging a real placenta, immersed in a tank filled with water, thanks to a fetoscope. The visual appearance of the 3D reconstruction obtained thanks to the proposed two-view Refractive Structure-from-Motion (second raw) and the two-view Structure-from-Motion are similar. However it can be observed a sensitive difference while comparing these reconstructions. It is more particularly the case in the vicinity of the veins visible at the surface of the placenta (last row)

algorithm allowed us to correct for these erroneous initialisation despite a loss in accuracy (see Fig. 11).

Toy dataset results related to this dataset are depicted in Figure 11. We observed similar accuracy using 2View SfM and 2View RSfM with a mean error of respectively 0.3 and 0.2 mm and a standard deviation of 0.4 and 0.5 mm. It can nevertheless be highlighted a slight improvement in uniformity of shape reconstruction which can be observed by looking at the distribution of the errors. Moreover, it is worth to note that our method allowed us to efficiently remove outliers which is essential in underwater imaging.

Fetoscopy dataset we aimed at reconstructing the surface of a human placenta immersed under water reproducing a real fetoscopic surgery scenario. Results are presented in Fig. 12. Quasi-dense 3D reconstruction in such case remains difficult due to the lack of features at the surface of the organ. We have however been able to obtain partial reconstruction for both 2View SfM and 2View RSfM thanks to the effectiveness of the proposed method for removing outliers. While 3D shape estimation appears visually similar for both 2View SfM and 2View RSfM, we observe significant disparities in the vicinity of veins and arteries. The proposed 2View RSfM method highlights more noticeably the slight changes at the surface of the placenta which is of particular importance in fetoscopy. We furthermore noticed that it was particularly significant when the viewing angle of the camera was important (side view of the placenta) which is consistent with the results obtained with the statuette dataset.

7 Conclusion

We have proposed a new formulation for the two-view geometry of images in underwater environments relying on the refractive epipolar geometry. We derived a novel rF constraint which forms the basis of a robust method for accurate and quasi-dense stereo reconstruction under water. More particularly, we showed that relative camera pose estimation based on the rF constraint outperforms state of the art techniques considered in underwater imaging. Numerical validation on synthetic data demonstrates the robustness of our approach against varying levels of additive noise while state of the art methods are particularly noise sensitive. This is an important improvement in underwater imaging context. We validated our method on real images by performing laboratory experiments and demonstrated that it improves underwater 2View SfM on both 3D shape and camera pose estimation. We demonstrated a potential application of 2View RSfM for endoscopic surgery. We successfully applied our method to fetoscopy and more particularly the 3D reconstruction of a placenta. Beyond this evaluation, the practical value of our work lies in various applications of underwater imaging where one can assume a single, thin refractive interface that separates the camera and the external environment.

Possible extensions of our work include implementing it within a more comprehensive pipeline for multiple-view reconstruction by extending our two view formulation for 2View RSfM. It would also be of interest to apply the con-

straint to develop a refractive Simultaneous Localisation and Mapping (SLAM) framework and potentially consider in more depth the relationship between the light source and the camera, which dictates the position of reflections within the scene's illumination. Theoretically our current assumption of a single refractive interface may also be extended to handle thick interfaces that are used for deep underwater reconstruction. Our method could then integrate motion information generally provided by underwater robotic system as an additional constraint.

Appendix A: Refractive Forward Projection Equation

Two constraints can be directly derived from the flat refractive camera model (Agrawal et al. 2012). They are illustrated in (Fig. 13) and recalled below in order to introduce the forward refractive projection equation.

The Flat Refraction Constraint (FRC) enforces the incident light ray $\mathbf{L}_j^i$ (i.e. travelling within the external medium) to be parallel to the line joining the point $\mathsf{R}_j \mathbf{Q}^i + \mathbf{t}_j$ and the incidence point $\mathbf{P}_j^i$. The rotation matrix R_j and the translation

Fig. 13 The refractive forward projection equation can be derived from the co-planarity constraint ensuring each light path lies on the plane of refraction (Agrawal et al. 2012)

vector $\mathbf{t}_j$ define the pose of the camera in view j.

$$(\mathsf{R}_j \mathbf{Q}^i + \mathbf{t}_j - \mathbf{P}_j^i) \times \mathbf{L}_{(1,2,3)}{}_j^i = 0 \quad (12)$$

The Co-planarity Constraint (CC) enforces each light ray to lie on the plane of refraction (defined by the camera axis and an incident light ray) and the refracted ray to also intersect the camera axis. This can be mathematically described by:

$$(\mathsf{R}_j \mathbf{Q}^i + \mathbf{t}_j)^\top (\mathbf{n} \times \mathbf{P}_j^i) = 0 \quad (13)$$

The refractive forward projection equation has originally been defined in Glaeser and Schröcker (2000) but it can also be efficiently derived from the CC (Agrawal et al. 2012). It allows us to compute the coordinates of the point of incidence $\mathbf{P}_j^i$ corresponding to the 3D point $\mathbf{Q}^i$. It is derived considering the 2D coordinate frame (u_1, u_2) of the corresponding plane of refraction, where the axis u_2 aligns with the camera's optical axis (i.e. $u_2 = n = Z$) and the axis $u_1 = u_2 \times (u_2 \times \mathbf{Q}^i)$. The point $\tilde{\mathbf{Q}}^i = (\tilde{\mathbf{Q}}_x{}^i, \tilde{\mathbf{Q}}_y{}^i)$ expresses the 3D point $\mathbf{Q}^i$ in this coordinate frame. Similarly, $\tilde{\mathbf{P}}_j^i = (\tilde{\mathbf{P}}_{xj}{}^i, \tilde{\mathbf{P}}_{yj}{}^i) = (\tilde{\mathbf{P}}_{xj}{}^i, d)$ corresponds to the point $\mathbf{P}_j^i$. The refracted light ray can then be expressed by $\tilde{\mathbf{P}}_{xj}{}^i u_2 + d u_1$. By deriving the CC in the coordinate frame of the plane of refraction, and considering a single and thin refractive interface, the refractive forward projection equation can be formulated by the following 4th degree polynomial equation in the unknown $\tilde{\mathbf{P}}_{xj}{}^i$ (Agrawal et al. 2012):

$$(\tilde{\mathbf{Q}}_x{}^i - \tilde{\mathbf{P}}_{xj}{}^i)^2 (d^2 \mu^2 + \mu^2 \tilde{\mathbf{P}}_{xj}{}^{i^2}) - (d \tilde{\mathbf{P}}_{xj}{}^i - \tilde{\mathbf{Q}}_y{}^i \tilde{\mathbf{P}}_{xj}{}^i)^2 = 0 \quad (14)$$

where μ corresponds to the refractive index of the external medium (e.g. water refractive index ≈ 1.33).

Appendix B: Single-view Refractive Geometry

For the sake of clarity, we here recall the refractive single-view geometry introduced in Chari and Sturm (2009).

An incidence point $\mathbf{P}_1^i$ lying on the refractive plane is expressed by:

$$\mathbf{P}_1^i = \left(-d \frac{p_{x1}^i}{p_{z1}^i} \quad -d \frac{p_{y1}^i}{p_{z1}^i} \quad 0 \quad 1 \right)^\top \quad (15)$$

reminding that $\mathcal{P}_j^i = (p_{xj}^i \ p_{yj}^i \ p_{zj}^i)^\top$ corresponds to the refracted light ray associated to the point $\mathbf{p}_j^i$ and d is the distance from the camera optical centre to the refractive plane.
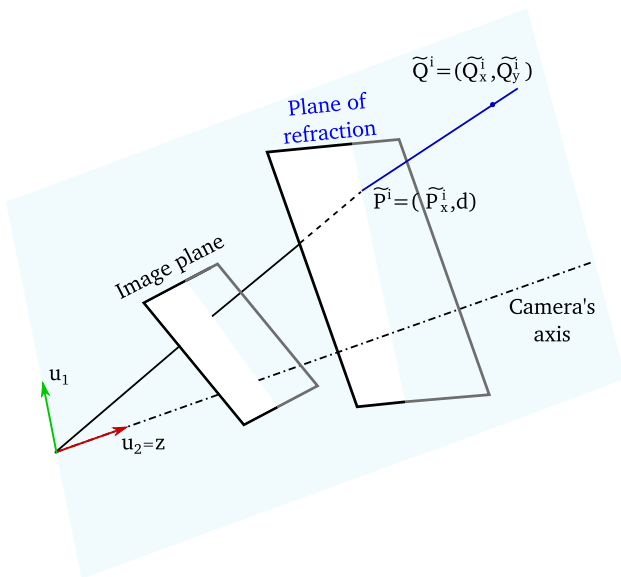
Snell's law (Hecht 1998) allows us to derive the direction vector of the corresponding incident light ray:

$$\mathcal{L}_1^i = (\lambda p_{x_1}^i \ \ \lambda p_{y_1}^i \ \ \sqrt{1 - \lambda^2 + \lambda^2 p_{z_1}^{i\,2}})^\top \quad (16)$$

where $\lambda = \frac{\mu_1}{\mu_2}$, μ_1 is the refractive index within the camera housing (generally equal to 1) and μ_2 is the refractive index of the refractive medium. We furthermore define $\gamma_{z_j}^i = \sqrt{1 - \lambda^2 + \lambda^2 p_{z_j}^{i\,2}}$.

Using Plücker coordinates, the incident light ray can therefore be formulated as:

$$\mathbf{L}_1^i = (\lambda p_{x_1}^i \ \ \lambda p_{y_1}^i \ \ \gamma_{z_1}^i \ \ -d\frac{p_{y_1}^i}{p_{z_1}^i}\gamma_{z_1}^i \ \ d\frac{p_{x_1}^i}{p_{z_1}^i}\gamma_{z_1}^i \ \ 0)^\top \quad (17)$$

The Plücker constraint on lines intersection directly derives from the definition of Plücker coordinates (Sturm and Barreto 2008). It allows us to express the intersection of the incident ray $\mathbf{L}_1^i$ with a 3D ray $\mathbf{L}'$ (i.e. a 3D ray or a 3D line $\mathcal{L}'$ intersecting the incident ray $\mathcal{L}_i^1$):

$$\mathbf{L}_1^{i\top} \mathsf{W}\mathbf{L}' = 0 \quad (18)$$

where $\mathsf{W} = \begin{pmatrix} 0 & \mathsf{I} \\ \mathsf{I} & 0 \end{pmatrix}$.

The previous relationship can be rewritten in the following form:

$$\lambda \mathbf{L}'^{\top}{}_{(4,5,3)_1}^{i} t_a \mathcal{P}_1^i + \frac{\gamma_{z_1}^i}{p_{z_1}^i} \mathbf{L}'^{\top}{}_{(6,1,2)_1}^{i} t_b \mathcal{P}_1^i = 0 \quad (19)$$

where $t_a = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ and $t_b = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -d & 0 \\ d & 0 & 0 \end{pmatrix}$.

Removing root square in $\gamma_{z_1}^i$ allows to reformulate equation (19) as follow (Chari and Sturm 2009).

$$\left(\lambda^2(-\hat{\mathsf{A}} + \hat{\mathsf{B}}) \ \ (1-\lambda^2)\hat{\mathsf{B}}\right) \underbrace{\begin{pmatrix} \mathsf{D}_t & 0 \\ 0 & \mathsf{D}_t \end{pmatrix}}_{\mathsf{D}'} (\widehat{\mathcal{P}_1^i} \ \ \frac{\widehat{\mathcal{P}_1^i}}{p_{z_1}^{i\,2}})^\top = 0 \quad (20)$$

where $\mathsf{A} = \lambda \mathbf{L}'^{\top}{}_{(4,5,3)_1}^{i} t_a$ and $\mathsf{B} = \mathbf{L}'^{\top}{}_{(6,1,2)_1}^{i} t_b$.

The refractive projection equation can then be derived in its general form by expending equation (20). Unlike the formulation proposed in Agrawal et al. (2012) (equation (14)), it does not directly depend on the planes of refraction.

$$\left(\widehat{\mathbf{L}}'^{\top}{}_{(6,1,2)_1}^{i} \ \ \widehat{\mathbf{L}}'^{\top}{}_{(4,5,3)_1}^{i}\right) \mathsf{rP}^\top \begin{pmatrix} \widehat{\mathcal{P}_1^i} \\ \frac{\widehat{\mathcal{P}_1^i}}{p_{z_1}^{i\,2}} \\ \widehat{\mathcal{P}_1^i} \end{pmatrix} = 0 \quad (21)$$

The refractive projection matrix rP is therefore defined in Chari and Sturm 2009 as:

$$\mathsf{rP} = \mathsf{D}'^\top \underbrace{\begin{pmatrix} (1-\lambda^2)\mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_b^\top \otimes \mathsf{t}_b^\top \mathsf{S}_t^\top & 0 \\ \lambda^2 \mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_b^\top \otimes \mathsf{t}_b^\top \mathsf{S}_t^\top & -\lambda^2 \mathsf{D}_t^{-1}\mathsf{S}_t \mathsf{t}_a^\top \otimes \mathsf{t}_a^\top \mathsf{S}_t^\top \end{pmatrix}}_{\mathsf{M}} \quad (22)$$

The refractive projection matrix is of size 12×12. It only depends on the medium refractive index on both sides of the refractive interface as well as on the distance between the optical centre of the camera and the refractive plane.

References

Agrawal, A., & Ramalingam, S. (2013). Single image calibration of multi-axial imaging systems. In *IEEE conference on computer vision and pattern recognition* (pp. 1399–1406). https://doi.org/10.1109/CVPR.2013.184.

Agrawal, A., Ramalingam, S., Taguchi, Y., & Chari, V. (2012) A theory of multi-layer flat refractive geometry. In *IEEE conference on computer vision and pattern recognition* (pp. 3346–3353). https://doi.org/10.1109/CVPR.2012.6248073.

Bouguet, J. Y. (2008). Camera calibration toolbox for Matlab.

Chadebecq, F., Vasconcelos, F., Dwyer, G., Lacher, R., Ourselin, S., Vercauteren, T., & Stoyanov, D. (2017). Refractive structure-from-motion through a flat refractive interface. In *The IEEE international conference on computer vision (ICCV)*.

Chang, Y. J., & Chen, T. (2011). Multi-view 3d reconstruction for scenes under the refractive plane with known vertical direction. In *IEEE conference on computer vision* (pp. 351–358).

Chari, V., & Sturm, P. (2009). Multiple-view geometry of the refractive plane. In *British machine vision conference* (pp. 1–11). https://doi.org/10.5244/C.23.56.

Chaudhury, S., Agarwal, T., & Maheshwari, P. (2015). Multiple view 3D reconstruction in water. In *National conference on computer vision, pattern recognition, image processing and graphics* (pp. 1–4). https://doi.org/10.1109/NCVPRIPG.2015.7490055.

Chen, X., & Yang, Y. (2014). Two-view camera housing parameters calibration for multi-layer flat refractive interface. In *IEEE conference on computer vision and pattern recognition* (pp. 524–531). https://doi.org/10.1109/CVPR.2014.74.

Daga, P., Chadebecq, F., Shakir, D., Herrera, L., Tella, M., Dwyer, G. (2016). Real-time mosaicing of fetoscopic videos using sift. In *International Society for Optics and Photonics of SPIE Medical Imaging* (pp. 97861R–97861R).

Fitzgibbon, A. W. (2001). Simultaneous linear estimation of multiple view geometry and lens distortion. In *IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. I–125).

Fryer, J., & Fraser, C. (1986). On the calibration of underwater cameras. *The Photogrammetric Record*, *12*(67), 73–85. https://doi.org/10.1111/j.1477-9730.1986.tb00539.x.

Gallup, D., Frahm, J. M., Mordohai, P., Yang, Q., & Pollefeys, M. (2007). Real-time plane-sweeping stereo with multiple sweeping directions. In *IEEE conference on computer vision and pattern recognition* (pp. 1–8). https://doi.org/10.1109/CVPR.2007.383245.

Glaeser, G., & Schröcker, H. P. (2000). Reflections on refractions. *Journal for Geometry and Graphics*, *4*(1), 1–18.

Grossberg, M., & Nayar, S. (2005). The raxel imaging model and ray-based calibration. *International Journal of Computer Vision*, *61*(2), 119–137. https://doi.org/10.1023/B:VISI.0000043754.56350.10.

Grossberg, M. D., % Nayar, S. K. (2001) A general imaging model and a method for finding its parameters. In *IEEE conference on computer vision* (Vol. 2, pp. 108–115). https://doi.org/10.1109/ICCV.2001.937611.

Haner, S., & Aström, K. (2015). Absolute pose for cameras under flat refractive interfaces. In *IEEE conference on computer vision and pattern recognition* (pp. 1428–1436). https://doi.org/10.1109/CVPR.2015.7298749.

Hartley, R., & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.

Hecht, E. (1998). Optics 4th edition. Addison Wesley Longman Inc.

Jordt, A. (2014). Underwater 3D reconstruction based on physical models for refraction and underwater light propagation. Ph.D. thesis.

Jordt, A., Köser, K., & Koch, R. (2016). Refractive 3D reconstruction on underwater images. *Methods in Oceanography*,. https://doi.org/10.1016/j.mio.2016.03.001.

Jordt-Sedlazeck, A., & Koch, R. (2012) Refractive calibration of underwater cameras. In *European conference on computer vision* (Vol. 7576, pp. 846–859). https://doi.org/10.1007/978-3-642-33715-4_61.

Jordt-Sedlazeck, A., & Koch, R. (2013). Refractive Structure-from-Motion on Underwater Images. In *IEEE conference on computer vision* (pp. 57–64). https://doi.org/10.1109/ICCV.2013.14.

Jordt-Sedlazeck, A., Jung, D., & Koch, R. (2013). Refractive plane sweep for underwater images. *Lecture Notes in Computer Science*, *8142*, 333–342. https://doi.org/10.1007/978-3-642-40602-7_36.

Kang, L., Wu, L., & Yang, Y. H. (2012a). Experimental study of the influence of refraction on underwater three-dimensional reconstruction using the svp camera model. *Applied Optics*, *51*(31), 7591–7603.

Kang, L., Wu, L., & Yang, Y. H. (2012b). Two-view underwater structure and motion for cameras under flat refractive interfaces. In *European conference on computer vision* (pp. 303–316).

Lavest, J., Rives, G., & Lapresté, J. (2000). Underwater camera calibration. *European Conference on Computer Vision*, *1843*, 654–668. https://doi.org/10.1007/3-540-45053-X_42.

Li, H., Hartley, R., & Kim, J. (2008). A linear approach to motion estimation using generalized camera models. In *IEEE conference on computer vision and pattern recognition* (pp. 1–8). https://doi.org/10.1109/CVPR.2008.4587545.

Lodi Rizzini, D., Kallasi, F., Oleari, F., & Caselli, S. (2015). Investigation of vision-based underwater object detection with multiple datasets. *International Journal of Advanced Robotic Systems*, *12*(77), 1–13. https://doi.org/10.5772/60526.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *IEEE conference on computer vision* (Vol. 2, pp. 1150–1157). https://doi.org/10.1109/ICCV.1999.790410.

Luczyński, T., Pfingsthorn, M., & Birk, A. (2017). The pinax-model for accurate and efficient refraction correction of underwater cameras in flat-pane housings. *Ocean Engineering*, *133*, 9–22. https://doi.org/10.1016/j.oceaneng.2017.01.029.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, *11*(2), 431–441.

Massot-Campos, M., & Oliver-Codina, G. (2015). Optical sensors and methods for underwater 3D reconstruction. *Sensors*, *15*(12), 31525–31557.

Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F., & Sayd, P. (2009). Generic and real-time structure from motion using local bundle adjustment. *Image and Vision Computing*, *27*(8), 1178–1193. https://doi.org/10.1016/j.imavis.2008.11.006.

Nikitichev, D., Shakir, D., Chadebecq, F., Tella, M., Deprest, J., Stoyanov, D., et al. (2017). Medical-grade sterilizable target for fluid-immersed fetoscope optical distortion calibration. *Journal of Visualized Experiments*, *120*, 1–8. https://doi.org/10.3791/55298.

Oleari, F., Kallasi, F., Lodi Rizzini, D., Aleotti, J., & Caselli, S. (2015). An Underwater stereo vision system: From design to deployment and dataset acquisition. In *Proceedings of of the IEEE/MTS OCEANS* (pp. 1–5). https://doi.org/10.1109/OCEANS-Genova.2015.7271529.

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

Sedlazeck, A., & Koch, R. (2012). Perspective and non-perspective camera models in underwater imaging, overview and error analysis. *Lecture Notes in Computer Science*, *7474*, 212–242. https://doi.org/10.1007/978-3-642-34091-8_10.

Shibata, A., Fujii, H., Yamashita, A., & Asama, H. (2015a). Absolute scale structure from motion using a refractive plate. In *IEEE/SICE international symposium on system integration* (pp. 540–545). https://doi.org/10.1109/SII.2015.7405037.

Shibata, A., Fujii, H., Yamashita, A., & Asama, H. (2015b). Scale-reconstructable structure from motion using refraction with a single camera. In *IEEE conference on robotics and automation* (pp. 5239–5244). https://doi.org/10.1109/ICRA.2015.7139929.

Steigman, S. A., Kunisaki, S. M., Wilkins-Haug, L., Takoudes, T. C., & Fauza, D. O. (2010). Optical properties of human amniotic fluid: implications for videofetoscopic surgery. *Fetal diagnosis and therapy*, *27*(2), 87–90.

Stoyanov, D., Marco, V. S., Pratt, P., & Yang, G. Z. (2010). Real-time stereo reconstruction in robotically assisted minimally invasive surgery. *Medical image computing and computer assisted,*. intervention.

Sturm, P. (2005). Multi-view geometry for general camera models. In *IEEE conference on computer vision and pattern recognition* (Vol. 1, pp. 206–212). https://doi.org/10.1109/CVPR.2005.237.

Sturm, P., & Barreto, J. (2008). General imaging geometry for central catadioptric cameras. In *European conference on computer vision* (pp. 609–622).

Sturm, P., & Ramalingam, S. (2004). A generic concept for camera calibration. *European Conference on Computer Vision*, *3022*, 1–13. https://doi.org/10.1007/978-3-540-24671-8_1.

Sturm, P., Ramalingam, S., & Lodha, S. (2005). On calibration, structure-from-motion and multi-view geometry for general camera models. In *International Society for Photogrammetry and Remote Sensing, Panoramic Photogrammetry Workshop* (Vol. XXXVI-5/W8)

Treibitz, T., Schechner, Y., Kunz, C., & Singh, H. (2012). Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(1), 51–65. https://doi.org/10.1109/TPAMI.2011.105.

Vedaldi, A., & Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms.

Yau, T., Gong, M., & Yang, Y. H. (2013). Underwater camera calibration using wavelength triangulation. In *IEEE conference on computer vision and pattern recognition* (pp. 2499–2506). https://doi.org/10.1109/CVPR.2013.323.

You, S., Tan, R., Kawakami, R., Mukaigawa, Y., & Ikeuchi, K. (2016). Waterdrop stereo. arXiv:1604.00730.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334. https://doi.org/10.1109/34.888718.